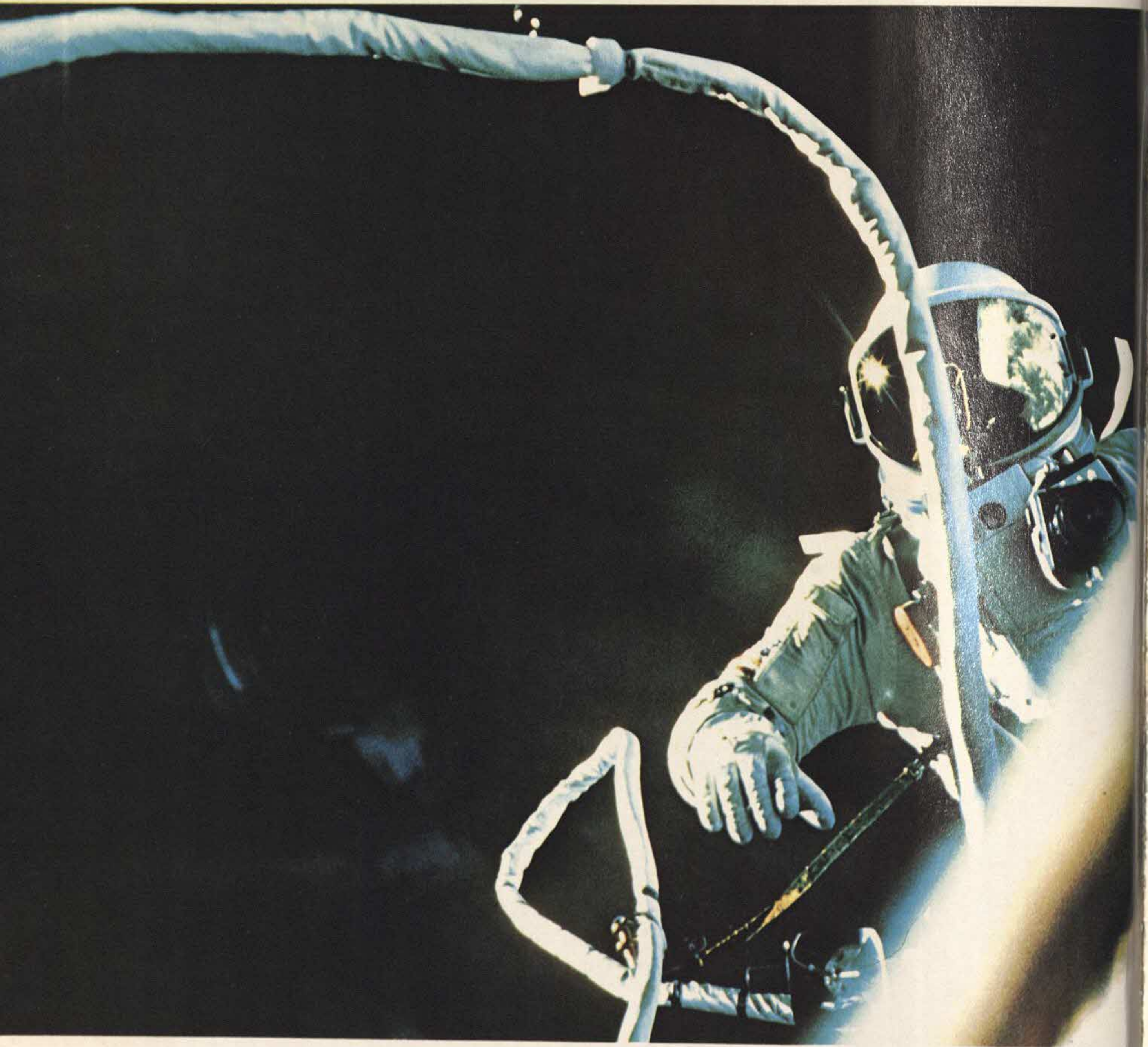
16

Illustrated World of Science Encyclopedia

Physics III:
ENERGY 2: LIGHT,
SOUND
AND
MECHANICS

THE WORLD OF SCIENCE



SPACE WALK — Because there is no atmosphere in space to reflect and diffuse sunlight, the sky appears black.

THE WORLD OF SCIENCE

VOLUME

16

PHYSICS III

Energy 2: Light, Sound, and Mechanics

with

The Illustrated Science Dictionary

CREATIVE WORLD PUBLICATIONS, INC.
CHICAGO

Copyright © 1971 by CREATIVE WORLD PUBLICATIONS, INC.
Illustrations copyright © 1966 by Fratelli Fabbri Editori.

THE WORLD OF SCIENCE ENCYCLOPAEDIA has been translated and adapted from SCIENZA: ENCICLOPEDIA TECNICA E SCIENTIFICA by Fratelli Fabbri Editori, Milan, Italy.

THE ILLUSTRATED SCIENCE DICTIONARY, copyright © 1971 by F. E. Compton Co., appears with full permission of the publisher, F. E. Compton Co., Chicago, Illinois.

All rights—including but without limitation copyright of text, drawings or any part of them, and the copyright of translations into any language—are reserved to the copyright proprietors.

No part of this work may be reproduced or utilized in any form by any means, electronic or mechanical—including but without limitation photocopying, recording, or by any information storage and retrieval system—without permission in writing from the copyright proprietors.

Library of Congress Catalog Card Number: 71-139118

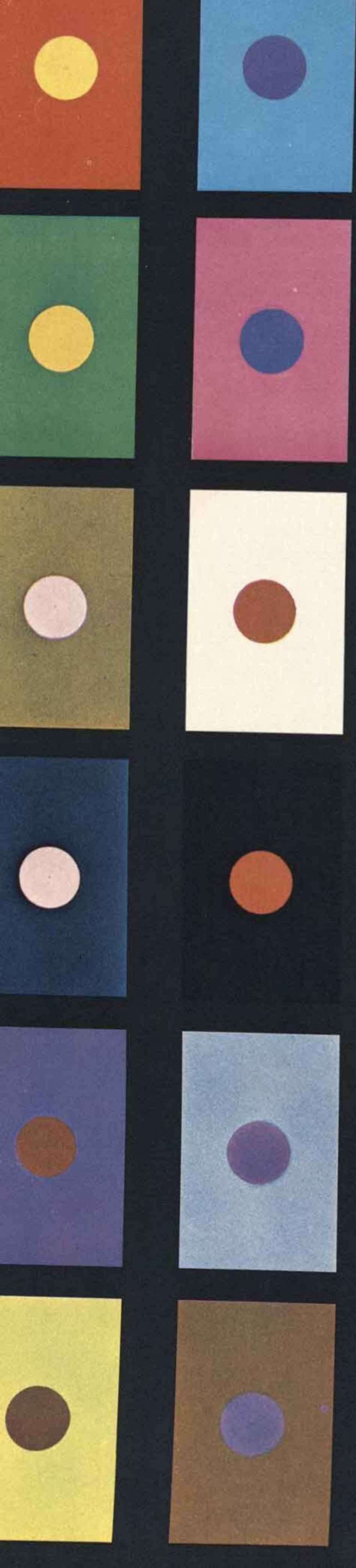
Typesetting primarily by SSPA Typesetting, Inc., Carmel, Indiana
Filmwork by Widen Engraving Co., Madison, Wisconsin
Printed and bound in U.S.A. by W. A. Krueger Co., Brookfield, Wisconsin

VOLUME 16

PHYSICS III

Energy 2 Light, Sound, and Mechanics

7 The Speed of Light
11 Reflection and Refraction
14 The Diffusion of Light
18 Phenomena of Light Diffusion
21 Light Absorption
24 Light Interference
27 Diffraction of Light
30 Resolving Power
33 Lenses and Aberrations
36 Dioptrics
40 Mirror Optics
43 Optical Illusions
46 Color
50 The Chromaticity Diagram
54 The Theory of Color
57 The Appearance of Color
60 Sources of Light
63 Excitons and Luminescences
66 Light and Crystals
68 The Intensity of Sound
70 The Characteristics of Sound
73 The Doppler Effect
76 Musical Acoustics
80 Architectural Acoustics
84 The Geometry of Freedom
87 Equilibrium
90 The Composition of Forces
93 The Equilibrium of Forces
96 Statics
99 Oscillating Motion
102 Composition of Motion
106 Displacement and Motion
110 Inertia
113 The Moment of Inertia
116 Impulse and Momentum
118 The Third Law of Motion
120 Friction
123 The Conservation of Energy
127 Power
130 Impact Phenomena
134 Hooke's Law
136 The Pendulum
140 Universal Gravitation
143 The Gradation of Forces
146 Gravitational Waves
150 Force Fields
155 Nuclear Reactions
157 The Production of Artificial Nuclei
161 *Illustrated Science Dictionary* (potential difference *to* reflux)



VOLUME 16

PHYSICS III

Energy 2: Light, Sound, and Mechanics

INTRODUCTION

Volume 16 examines many of the physical phenomena associated with light, sound, and mechanics.

Light, or optics, is the study of those electromagnetic waves to which the human eye is sensitive. Their wavelengths lie in the very narrow band between 4×10^{-5} cm (violet) and 7×10^{-5} cm (red). Because of their intimate relationship with everyday experiences, the effects of light have always provided a fascination for man. Who has not wondered what causes a rainbow, or why the sky is blue by day and red at sunset? Today physicists can explain, through the basic laws of light, such phenomena as reflection, refraction (the bending of light by a prism), and the coloring of an oil film on water. Throughout the centuries scientists have made great use of these properties in designing optical instruments to aid their investigations of natural phenomena. Today it is possible to build (often with the aid of a computer) complicated lens systems for microscopes, telescopes, and cameras.

Light has another interesting aspect. Electromagnetic radiation in the optical band originates from transitions among the outer electrons of atoms and molecules; hence, in order to explain some of the phenomena associated with light, it is necessary to know atomic and molecular structures. The effect of light on outer electrons is crucial to life on Earth; the sun's rays bring about many of the subtle chemical processes (such as photosynthesis) that are often taken for granted but on which life depends. Knowledge of atomic properties has recently been put to use in the development of the laser, an extremely high-intensity beam of monochromatic light.

The branch of physics known as sound concerns the sense of hearing. As in the case of light, sound has a wave nature. Usually, many frequencies are present in a given sound, and if the frequencies are in simple ratios the sound is usually considered pleasing. Musical instruments are therefore constructed to produce waves of the appropriate frequencies. This can be done in many ways; for example, by causing vibrations of a stretched wire (violin, piano), or of a column of air (flute, clarinet, or trumpet), or of a stretched membrane (drum).

The third branch of physics considered in this volume is mechanics, the study of objects in motion. The laws of classical mechanics were first formulated more than three hundred years ago by Newton and Galileo, and the subject has continued to develop since then. More recently, powerful analytic methods have been introduced; Einstein's theory of relativity has been incorporated; and in the 1920s quantum mechanics was invented to give a more correct description of the structure of atoms.

Mechanics holds a special role in physics because it is the subject that contains many of the basic concepts that are convenient to use in most areas. The meaning of force, potential and kinetic energies, and momentum are just as applicable in electromagnetism and in atomic and nuclear physics as they are in mechanics.

The law of universal gravitation is another vital contribution by Isaac Newton to science. This law explains how the moon and the planets are held in their path by exactly the same reasoning that explains why a ball drops at the Earth's surface. Newton's law of gravitation helped explain many interesting phenomena observed in the skies and later led to the predictions of the location of a new planet, Neptune.

ARCHIBALD W. HENDRY, Ph.D.
Assistant Professor of Physics,
Indiana University

THE SPEED OF LIGHT

from the Earth to the moon in little more than a second

In ancient times, scientists considered the propagation of light to be instantaneous—that light arrived at all points of space at the same instant it was emitted from a source. Their error is understandable because the speed of light is enormously great and it is difficult to think of a simple experiment to prove their conviction wrong.

As a comparison, the speed of sound is also great—about 0.3 km/sec (about 1,000 ft/sec); however, many natural phenomena enable us to realize that its speed is finite. The ancients—as well as modern men—could observe the flash of lightning before they heard the thunderclap, and see the fall of a heavy object at some distance before they heard its crash.

Light, however, has a speed so great that it is difficult to realize that it, too, has a finite value. The dimensions of the largest laboratory are too small to determine that light is not propagated over an infinite distance in a finite period of time—unless great ingenuity or extremely sophisticated instruments are used. To the early scientists, even the dimensions of the Earth were insufficient to prove or disprove their theories. It is now known that light is propagated at a speed of about 300,000 km (about 186,000 mi) per second—seven times greater than the Earth's circumference. The distance sound travels in one second is covered by light in one millionth of a second. Light travels from the moon to the Earth in little more than one second; it travels from a house lamp to the wall of the room in less than ten billionths of a second; it travels a distance equal to the diameter of a man's thumb in a tenth of a billionth of a second—and today's physicists with modern equipment can measure its speed over such a short distance with considerable accuracy.

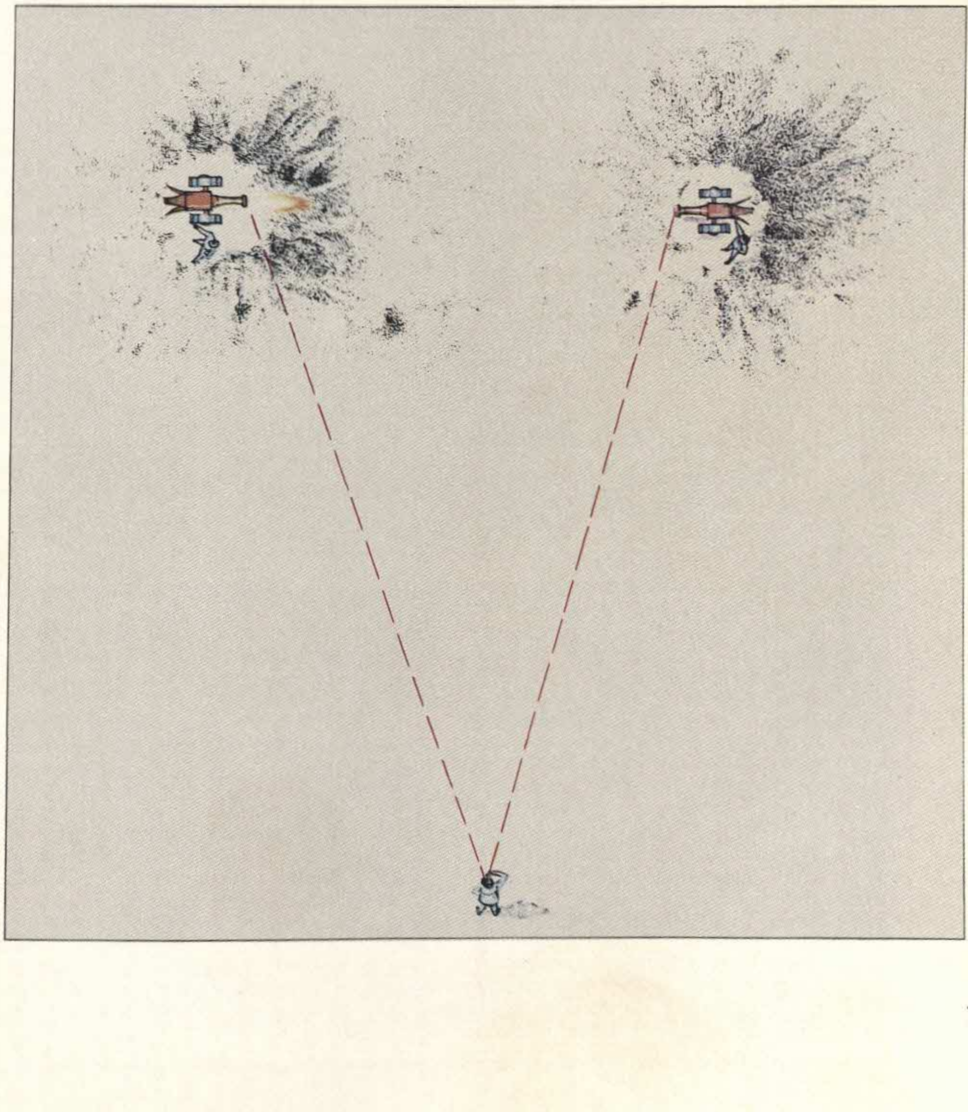
Early experiments regarding the propagation of light were conducted for the sole purpose of demonstrating that it was not propagated instantaneously, but at a finite speed. Not until later were experiments conducted to determine the actual speed of propagation.

Knowledge of the speed of light is of fundamental importance today and is largely used in practical ways. As an example, Roemer's measurement of the diameter of the Earth's orbit (explained in Illustration 2), using the speed of light, is one of the most accurate that can be obtained. The speed of light is basic to the use of radar (*ra*dio *d*etecting *a*nd *r*anging). All radar installations measure distances to objects by determining how long microwaves take to cover the distance to and from the target (radio waves travel at the same rate as light).

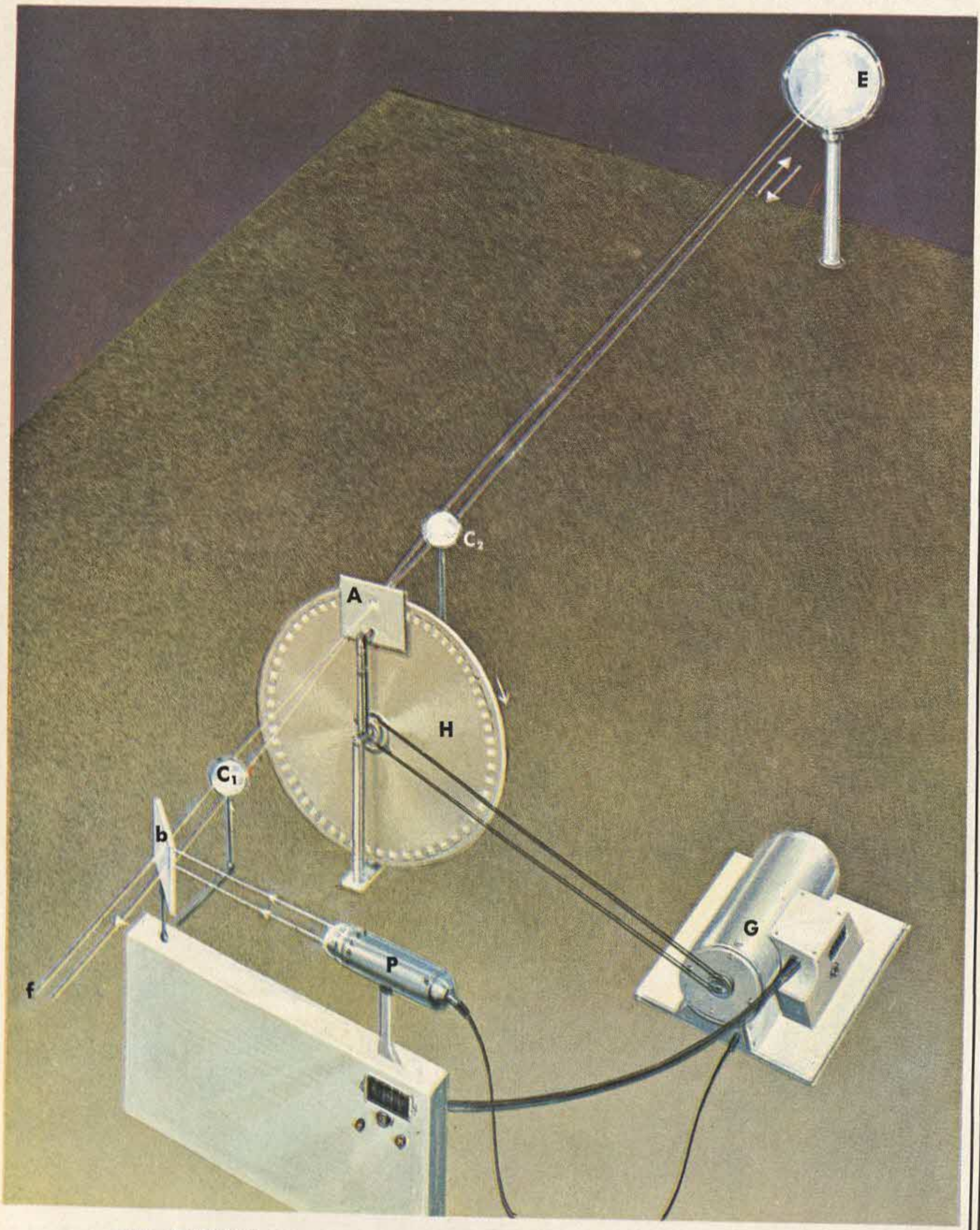
GALILEO'S ATTEMPT—The Italian astronomer and physicist Galileo devised an ingenious experiment for the purpose of determining whether the propagation of light was instantaneous, or whether it occurred at a finite speed. This was the first attempt to measure the speed of light. Although Galileo's experiment did not solve the problem, it is an interesting example of the difficulties encountered in seventeenth-century technology.

Galileo placed two cannons a few kilometers apart, on the slopes of two neighboring hills. Taking a position at a point equidistant from the two cannons, he was able to view both weapons simultaneously. The gunner on one cannon was ordered to fire first; the other gunner was directed to fire only when he saw the flash of the first cannon. Galileo recognized that if the second gunner had infinitely fast reflexes—and light was propagated instantly—he would see the two cannons fire simultaneously. If the second gunner (under the same conditions) fired later than the first, Galileo would be able to notice the difference in time between the flashes and be able to estimate the speed of light. In practice, of course, the second gunner always fired after the first and it was obvious that his delay was caused by the time he needed to react. Galileo concluded that either the speed of light was truly infinite, or that it was too great to be measured by the methods available. He limited himself to stating that the speed of light must be many times that of the speed of sound.

1



2



THE PRINCIPLE ON WHICH MODERN MEASUREMENTS ARE BASED—Before the beginning of the present century, the French physicists Armand Fizeau and Léon Foucault managed to determine the speed of light, not by measuring the time it takes to cover a great distance (such as the Earth's diameter), but by measuring the time it takes to travel a few kilometers—or even a few meters within a laboratory. They based their experiments on a principle that has been used by other scientists of this century, using more modern equipment. The illustration shows the manner in which the original experiment was conducted. Light emitted by projector **P** is reflected by the semitransparent mirror **b** and the lens C_1 and is made to converge on slit **A**. The rays of light are able to continue on their way only if—after passing through **A**—they align with one of the holes cut into the circumference of the disk **H**. These holes are equidistant from each other. The mirror **E** reflects the beam of light, and it is converged again by lens C_2 into **A**. If the disk **H** is not in movement and has a hole aligned with slit **A**, the light will pass through and be collected at **F**.

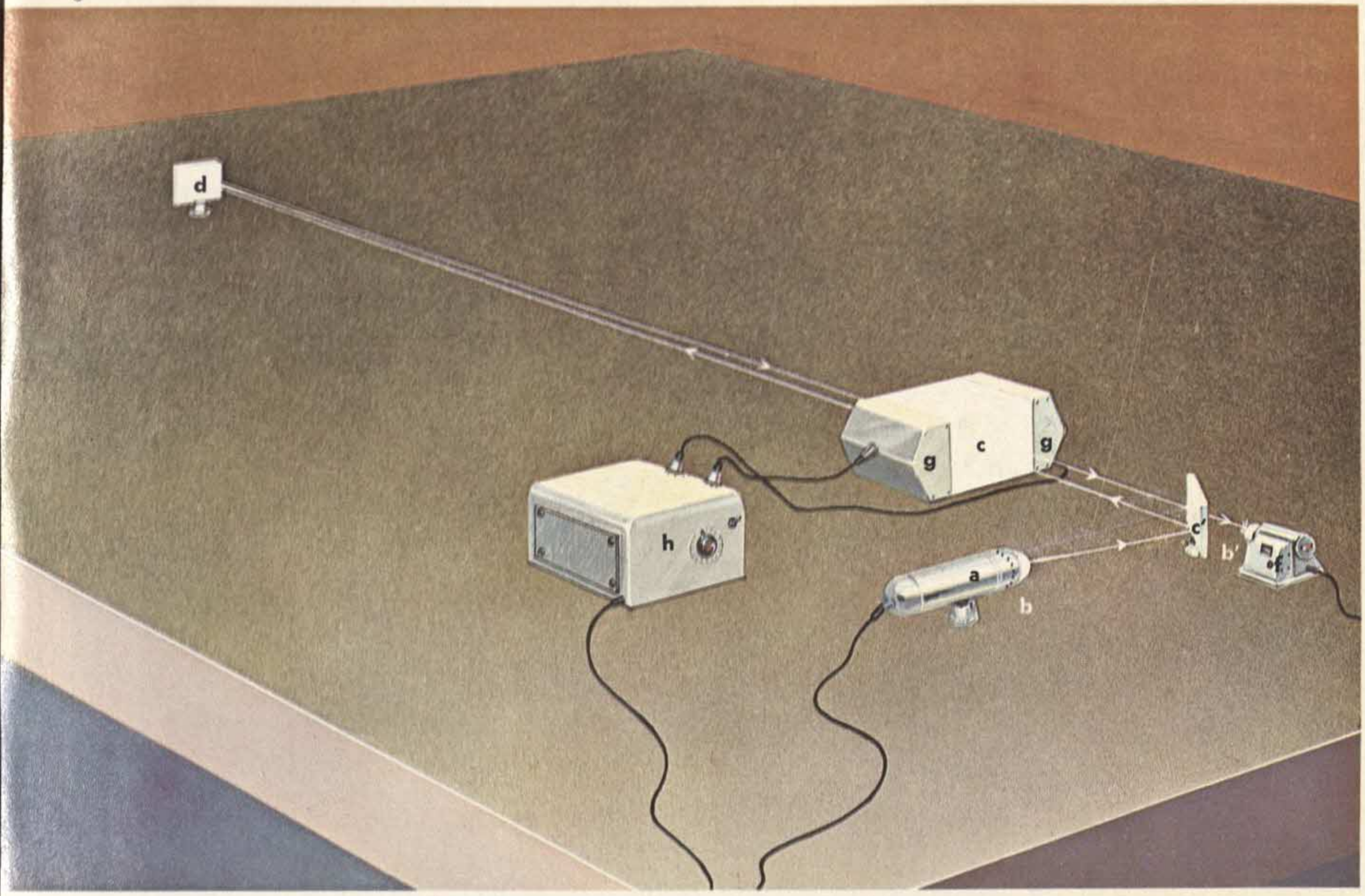
When the disk is rotated and its speed gradually increased, the light from **E** will initially strike solid parts of the disk. As the speed is increased, conditions can be created where the light from **E** begins to find the next hole aligned with slit **A**. The speed of the disk can be measured; the number of holes in the disk are known, and the distance between **A** and **E** can be determined. With these factors known, the speed of the light ray can be calculated. The speed is given by the distance the light travels during its two passages through the disk (from **A** to **E** and return), divided by the time taken to travel this distance, or **v = 2AE/t.**

MEASUREMENT BY MEANS OF KERR CELLS—The experimental layout shown in this illustration is an example of how the principle employed by Fizeau and Foucault (Illustration 3) is used with modern instruments. The light from source a passes through lens **b**, which turns it into a beam. The beam is directed to **c′**—a half-silvered glass—and deflected to the left. It then passes through **c**, a tank filled with nitrobenzene. This liquid, by exploiting the properties of polarized light, permits or prevents the passage of light, according to whether or not the tank is surrounded by an electric field. If the light passes through, it continues to mirror **d** and returns by its original path to the tank **c**. When the light encounters the tank on its return trip it will pass through it—if the electric field has not had sufficient time to change polarity. (If the polarity has changed, the light will not pass through.) The light, on passing through the tank, continues to the semitransparent mirror **c′**, passes through it to lens **b′**, and is concentrated by the lens onto a sensitive photoelectric cell **f.**

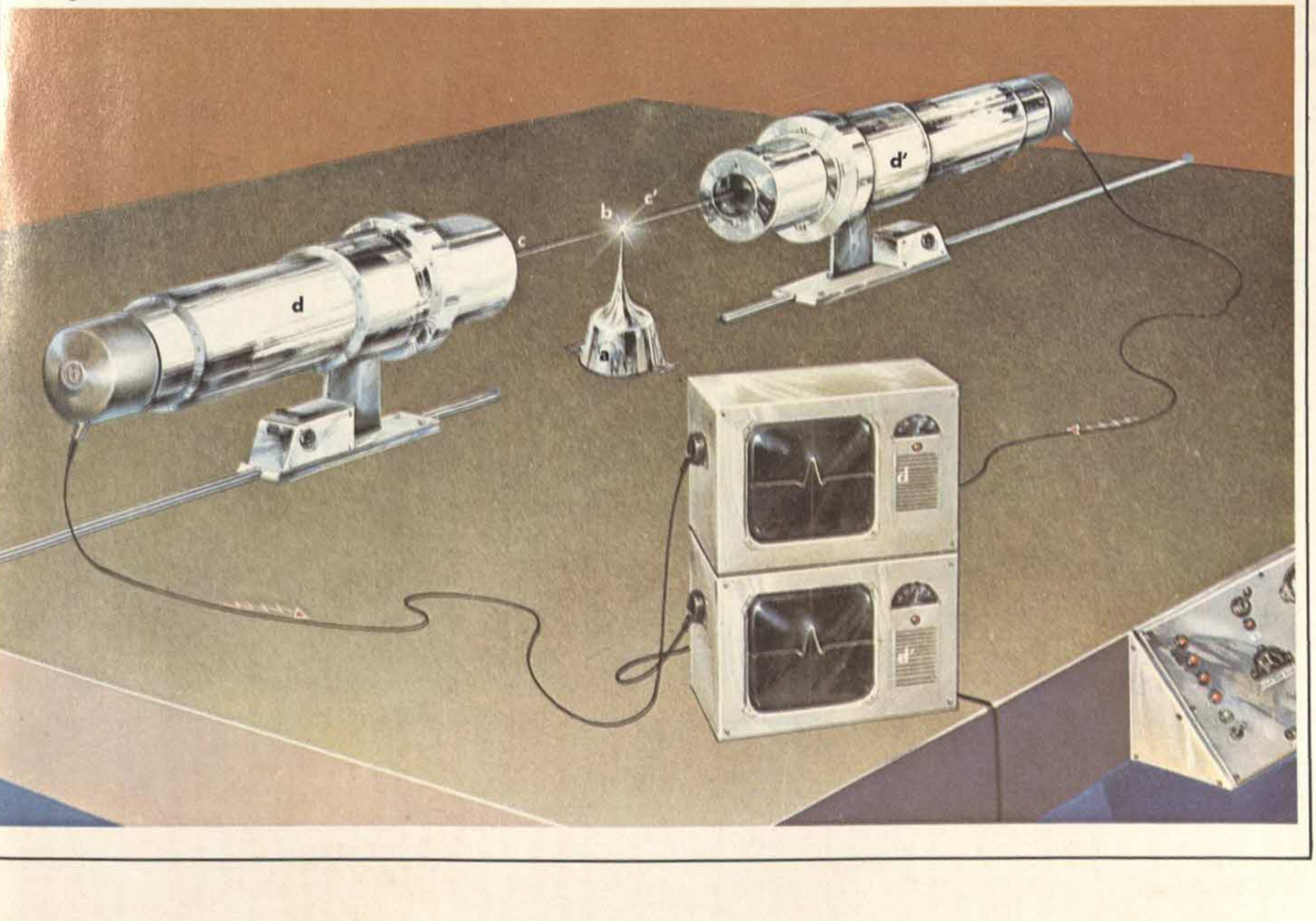
The tank **c** is called a Kerr cell, named after the Scottish physicist John Kerr, who discovered the property that enables nitrobenzene to be used as a shutter. The Kerr cell is subjected to a rapidly changing electric field by means of two metal plates **g.** These plates are connected to a generator producing a high-frequency alternating current. The frequency can be varied until the photoelectric cell no longer perceives light arriving from mirror **d.** The time taken by the light to travel twice the distance between the photoelectric cell and mirror **d** is exactly the time needed by the electric field of the Kerr cell to invert its polarity. Measuring this time and the distance between the photoelectric cell and the mirror gives the value of the speed of light.

THE DE BENEDETTI-RICHINGS EXPERIMENT—In 1954, two American physicists—De Benedetti and Richings—determined the time that light needs to travel as small a distance as 3 cm. Their experiment did not use light at all; it was done with gamma rays. Gamma rays are radiations of the same type as light rays, but with a wavelength several thousand times shorter than that of light. The method used by the two scientists is shown in the illustration. In the center **a** is a source of positive electrons with a diameter of about 1 mm (about 0.04 in.). Positive electrons (positrons) are a form of antimatter. When positrons encounter the negative electrons of ordinary matter **b**, the two particles spontaneously annihilate and turn into radiation, resulting in pairs of gamma rays that move off in opposite directions **c** and **c′**. At a short distance from the central source of positrons, the two scientists placed the apparatuses labeled **d** and **d′**; each of these was capable of emitting an electric pulse when crossed by a gamma ray. With **b** equidistant from **d** and **d′**, the gamma rays arrived simultaneously at the two detectors. If source **a** was displaced toward one of the two detectors (even if it was moved no more than 1.5 cm), the two pulses were generated at an interval of 10^{-10} seconds. With the two distances between the detectors and the source differing by 3 cm, it is possible to deduce the speed of light as 3×10^5 km/sec—or 300,000 km/sec. The De Benedetti-Richings experiment, although it produced a measurement of the speed of light, was not intended to make an accurate measurement; rather, it served to demonstrate the possibility of measuring the extremely brief time interval that elapses as light travels a few centimeters.

3

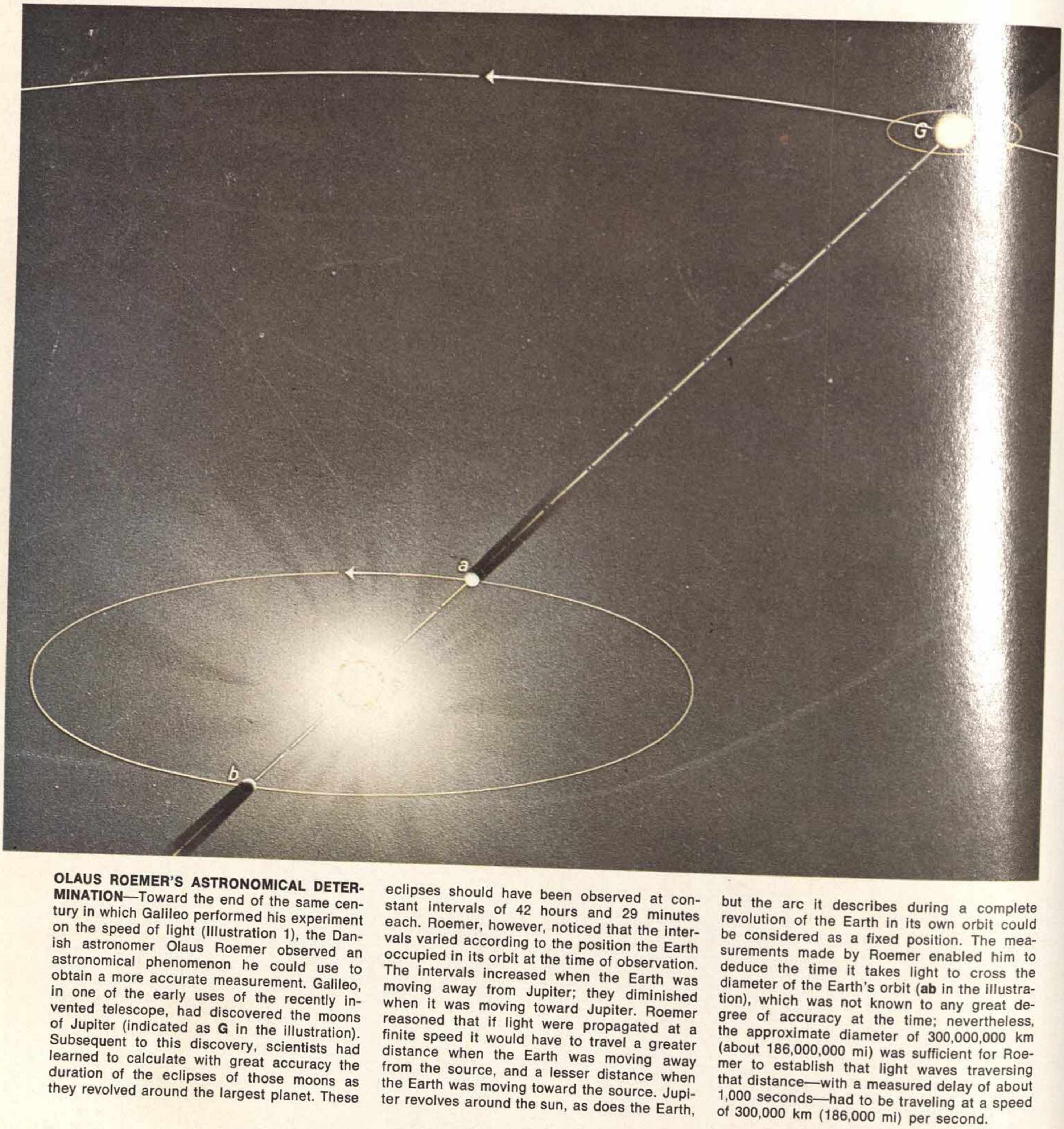


4



Refined knowledge of the speed of light now permits the most delicate measurements of distance. Geographical maps can be prepared using the same principle employed in radar, but over minute distances. All these developments are the product of centuries of scientific effort devoted to understanding and measuring the velocity of light.

5



OLAUS ROEMER'S ASTRONOMICAL DETERMINATION—Toward the end of the same century in which Galileo performed his experiment on the speed of light (Illustration 1), the Danish astronomer Olaus Roemer observed an astronomical phenomenon he could use to obtain a more accurate measurement. Galileo, in one of the early uses of the recently invented telescope, had discovered the moons of Jupiter (indicated as **G** in the illustration). Subsequent to this discovery, scientists had learned to calculate with great accuracy the duration of the eclipses of those moons as they revolved around the largest planet. These eclipses should have been observed at constant intervals of 42 hours and 29 minutes each. Roemer, however, noticed that the intervals varied according to the position the Earth occupied in its orbit at the time of observation. The intervals increased when the Earth was moving away from Jupiter; they diminished when it was moving toward Jupiter. Roemer reasoned that if light were propagated at a finite speed it would have to travel a greater distance when the Earth was moving away from the source, and a lesser distance when the Earth was moving toward the source. Jupiter revolves around the sun, as does the Earth, but the arc it describes during a complete revolution of the Earth in its own orbit could be considered as a fixed position. The measurements made by Roemer enabled him to deduce the time it takes light to cross the diameter of the Earth's orbit (**ab** in the illustration), which was not known to any great degree of accuracy at the time; nevertheless, the approximate diameter of 300,000,000 km (about 186,000,000 mi) was sufficient for Roemer to establish that light waves traversing that distance—with a measured delay of about 1,000 seconds—had to be traveling at a speed of 300,000 km (186,000 mi) per second.

REFLECTION AND REFRACTION

A telescope, a microscope, and a pair of binoculars are but a few of the instruments used to combine the laws of reflection and refraction to extend human vision. The scope of the eye can be expanded a thousandfold by these seemingly simple optical devices. A thorough examination of each device would reveal only two types of light-bending optical parts. The reflective surfaces of mirrors cause an abrupt change in the direction of a light ray, while the partial bending at the surfaces of lenses and other refractive parts is more subtle. However, all optical instruments employ the same optical parts, which differ only in design and location within the instrument.

The laws of reflection and refraction of light are fundamental to a study of optical instruments, and these laws represent the foundation for the study of light itself. A thorough knowledge of reflection allows the astronomer to design a curved mirror that can gather the light from a distant star for his examination. The lensmaker can draw on his knowledge of refraction to design lenses for microscopes, telescopes, or eyeglasses. The study of optical phenomena is no longer limited to visible light. Radar employs the laws of reflection to locate distant thunderstorms with radio waves in much the same way that an image is located in a plane mirror. Radar meteorology is the study of radar echoes received from natural reflecting sources in the atmosphere and the way the echo characteristics relate to the atmospheric processes influencing the reflecting mediums. Natural reflecting sources include hydrometeors, regions of perturbed refractive index, and paths of lightning discharges.

1

SOURCES OF LIGHT—Anything that can be seen can generally be defined as a source of light. The printed page, the electric lamp (Illustration 1a), and the candle flame (Illustration 1b) are all sources of light. A more rigorous definition would include all objects that emit any form of electromagnetic radiation. (Visible light is but a small portion of the electromagnetic spectrum.)

Although the printed page and the candle flame both qualify as sources of light, they differ considerably in the method by which they emit light. The printed page must be illuminated by light of another source before it can emit light of its own, and is, therefore, defined as a secondary source or a passive emitter. The candle flame, the electric lamp, and the sun are examples of primary light sources or active emitters. Active emitters of all types are considered luminous. Light is always emitted in the same way whether the source is active or passive.

Other characteristics, such as color and intensity, are often quite important to the observer. Most observations of optical phenomena require quite intense, small light sources. Such an emitter is called a point source.

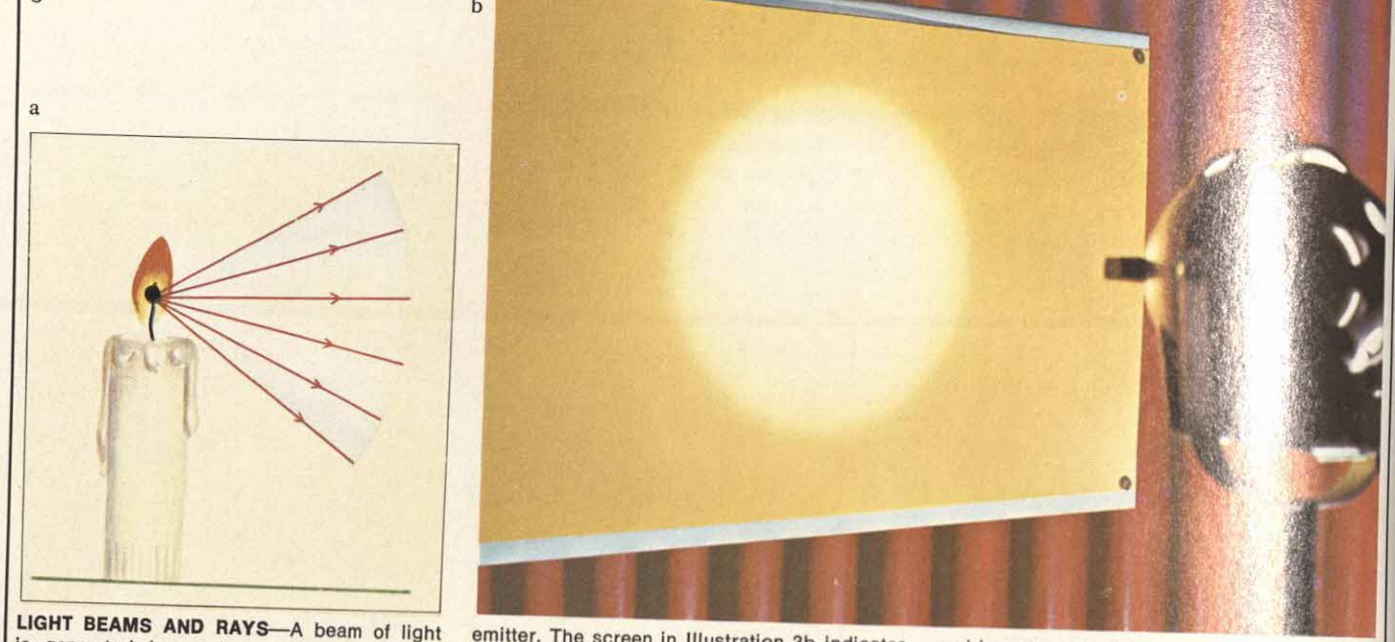
a

b

2

MOST SOURCES ARE NONDIRECTIONAL—A candle placed near a sheet of paper or in a box will illuminate equally in all directions. The light from the candle in the illustration cannot penetrate opaque objects, and an absence of light occurs at the base of the candle as well as outside the box. However, if the candle is held in a horizontal position, the light will fall uniformly on the bottom of the box, indicating that light does radiate equally in all directions. Further experimentation would show that the light decreases in intensity in all directions as the distance from the source increases. The example of the candle can be expanded to include all ordinary sources of light.

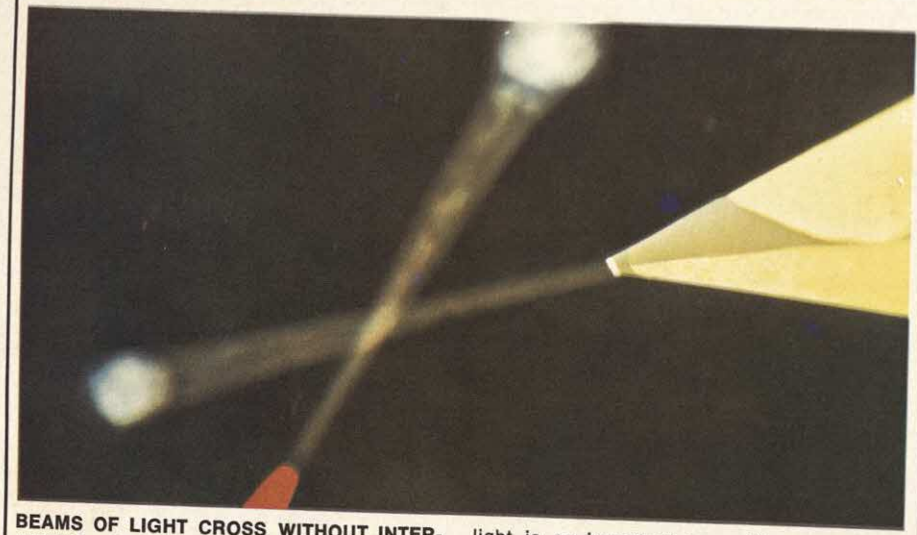
Because the laser is not an ordinary source of light, it is an exception to this rule. The laser is designed to emit light in only one direction. Laser light is produced by a highly controlled atomic mechanism and is emitted as a nearly cylindrical pencil of light.



LIGHT BEAMS AND RAYS—A beam of light is generated by enclosing the source and allowing the light to pass through a small opening in one direction. Because it is not an emitter, the beam cannot be seen. An object placed in the beam of light will interrupt the beam and become visible as a secondary emitter. The screen in Illustration 3b indicates the presence of the beam. A beam of light may be considered a bundle of straight lines radiating from the source. (These straight lines are used to explain light and are not a property of the light itself.)

The ray of light is an extremely small beam and is represented by one of the lines in Illustration 3a. All rays of light from a point source intersect at the source. The line or line segment that represents a ray can be described as the axis of a small cylindrical beam of light. Experiments involving optical phenomena describe light as a ray.

4

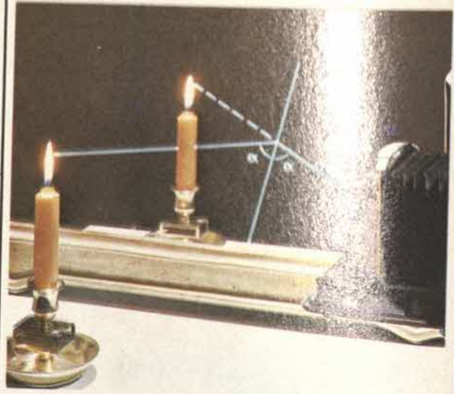


BEAMS OF LIGHT CROSS WITHOUT INTERACTING—Experiments carried out in the laboratory, on the surface of the Earth, or even on an astronomical scale indicate that light is propagated in straight lines. The endless applications of this phenomenon include surveying, building construction, astronomy, and, fundamental to all, vision. The eye is capable of reconstructing images only because light travels in straight lines. Scientists use the term rectilinear propagation to describe the straight-line nature of light. Only the most complex experiments would show any deviation of light from its path.

The illustration shows two beams of light crossing in a smoke-filled box. It can be seen that the two beams of light pass through each other without any interaction. This property of light is as important as rectilinear propagation. If light beams or rays could interfere with each other, vision would be impossible; only point sources of light in completely dark rooms could be seen, and even then only if no other objects were present to reflect the light.

In this example the two sources of light are completely incoherent and are uncorrelated with one another. However, in the case where light beams are coherent, interference between them takes place. An example of interference can be seen when a point source of light (or a line source) is placed behind an opaque sheet with two parallel slits cut in it. The two light beams emerging from the sheet will interfere with each other; the interference exhibits itself in alternating regions of light and darkness.

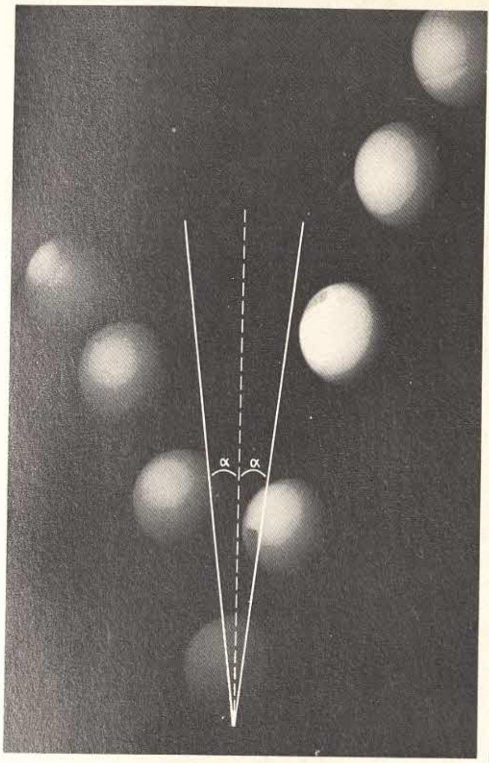
5



REFLECTION—A ray of light continues as a straight line from the source toward infinity unless it is stopped or deviated by an obstacle. On striking an obstacle, the ray may be totally absorbed and hence terminated, or it may be reflected. Most practical examples indicate that some light is reflected and some is absorbed. The nature of the obstacle determines the type of reflection. Diffuse reflection or scattering of the light occurs when light strikes an irregular surface such as a sheet of paper. Specular reflection occurs at smooth and regular surfaces. The mirror in the illustration produces specular reflection. The ray of light continues in a straight line with only its direction changed. Any surface that is capable of causing a specular reflection is defined as a specular surface.

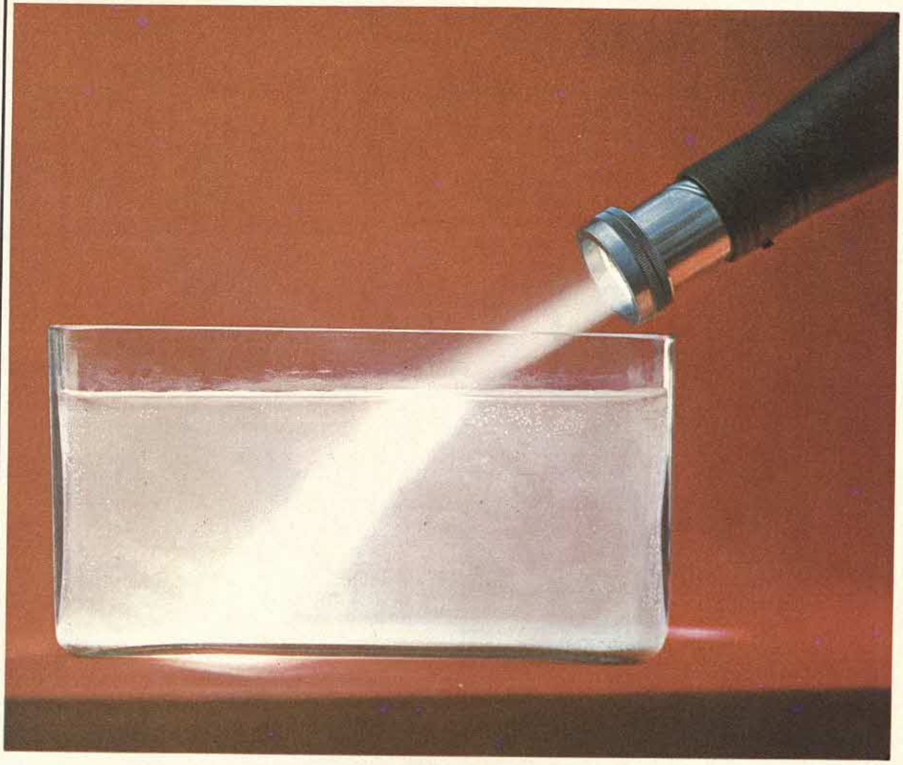
The incident ray (incoming) and the reflected ray (emerging) always form angles with the normal to the surface. A line perpendicular to the surface at the point of reflection is defined as the normal line. It should also be noted that the incident ray, the reflected ray, and the normal line always lie in the same plane.

6



A BOUNCING BALL—A ball bouncing from a smooth flat surface forms an incident angle that is equal to the reflected angle. The dotted line represents the normal to the surface at the point of reflection, and the Greek letter alpha (α) is used to show both the angle of incidence and the angle of reflection. The reflection of a ray of light from a specular surface can be likened to the mechanical example of the bouncing ball. (This example does not indicate that light always behaves as a bouncing ball.)

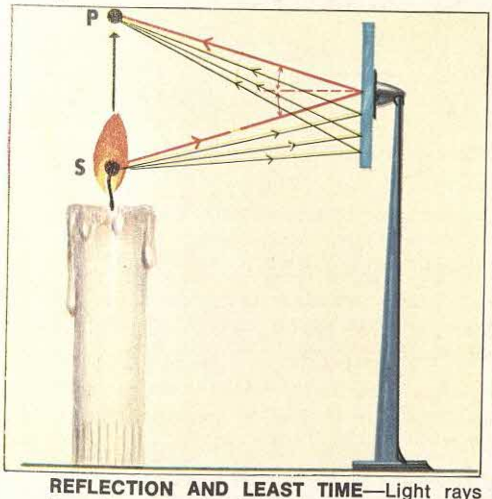
7



REFRACTION—If a ray of light is allowed to strike the surface of a liquid at an angle, the phenomenon of refraction is indicated by the bending of the light ray. As the ray of light passes from air to water, as in the illustration, it forms an angle that is nearer the perpendicular in the water than in the air. The degree of refraction observed with a substance is an inherent property of the substance, and a mathematical relationship can be drawn to describe this phenomenon. Refraction occurs at any surface between two different substances, providing a ray of light can pass between them. The closer the incident ray approaches the normal, the less the degree of refraction, while the greater the angle of incidence, the greater the degree of refraction.

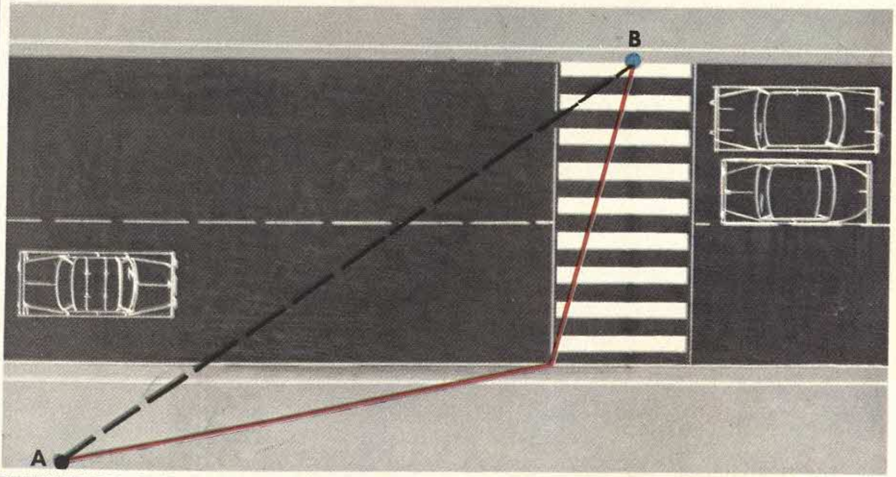
The path that light follows is reversible, and this property is useful in discussing refraction. Light generated within the medium is refracted as it emerges from the surface. A pond, a lake, a swimming pool, or even a glass of water appears to be shallower to the observer due to refraction of the light emerging from the surface.

8



REFLECTION AND LEAST TIME—Light rays from the candle strike the surface of a mirror. All rays originate at the source **S,** but only the red line strikes the point **P.** The incident ray and the reflected ray must form equal angles with the line normal to the surface, and only the ray represented by the red line does so. The shortest distance from source to point, this path represents the least time required for the light from the candle to be reflected to the point.

9

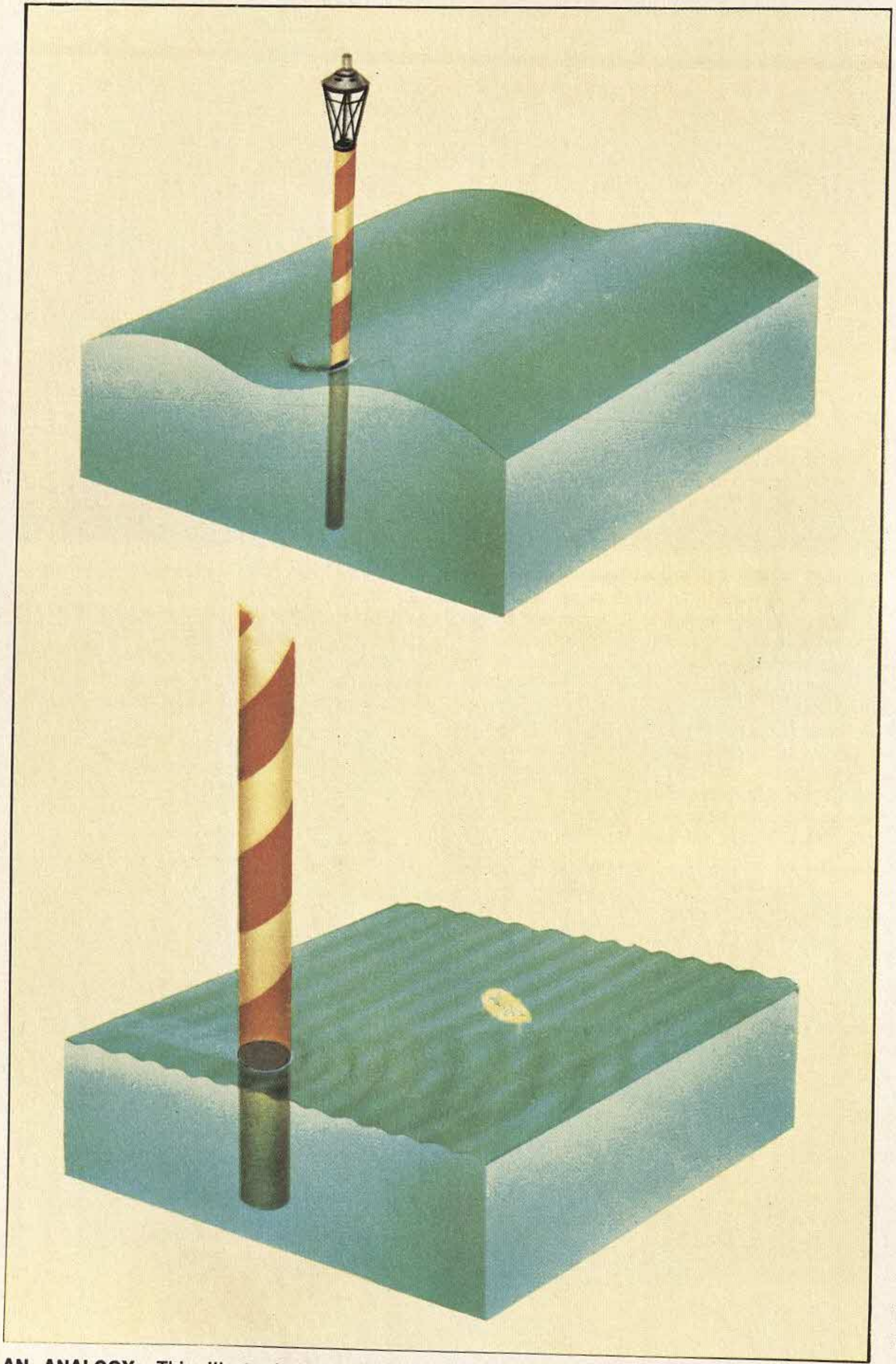


REFRACTION AND LEAST TIME—The principle of least time may be expanded to include an elementary example of refraction. The bending of a ray of light, as it passes from one medium to another, is caused by the different speeds of light in the substances. The speed of light in air is about 3×10^8 m/sec (about 186,000 mi/sec), while in water it is only about 2.25×10^8 m/sec (about 135,000 mi/sec). The degree of refraction is dependent on the differences in the speeds of light in the mediums.

The illustration likens the speed of a person walking to the speed of light. The greatest speed can be accomplished while walking on sidewalks; the speed is slowed considerably in the street. A person at **point A** wishing to cross the street and reach point **B** would choose to follow the red line. The red line would represent the minimum time to cross the street even though the path is longer if the walking speeds on the sidewalk and in the street are considered.

THE DIFFUSION OF LIGHT | the phenomenon of scattering

1



AN ANALOGY—This illustration attempts to show how waves behave when they encounter an obstacle. Two extreme cases are considered. In the first case (top illustration), a wave of considerable proportions (height or amplitude and wavelength) encounters a pole immersed in the water; the wave simply moves past the obstacle on both sides, and only a small part of the wave is reflected by the pole. Careful observation of this phenomenon reveals a small circular wave spreading out from the obstacle. In the second case (bottom illustration), a much smaller wave strikes the pole, and almost all of the wave is reflected. The disturbance is propagated away from the obstacle for a considerable distance. This phenomenon is somewhat analogous to that occurring when light waves strike an obstacle—a grain of dust, for example, or a droplet of water in a cloud or fogbank—an obstacle whose dimensions are approximately equal to the wavelength of the radiation.

Countless persons have stood on the seashore and watched the waves battering various obstacles—rocks, poles, and the shore itself. These observers have noticed how the obstacles at times stop the waves and at other times reflect them or drive them back. The behavior of ocean waves is somewhat analogous to that of light waves. When light waves encounter obstacles, they too may be absorbed, reflected, or diffused (scattered).

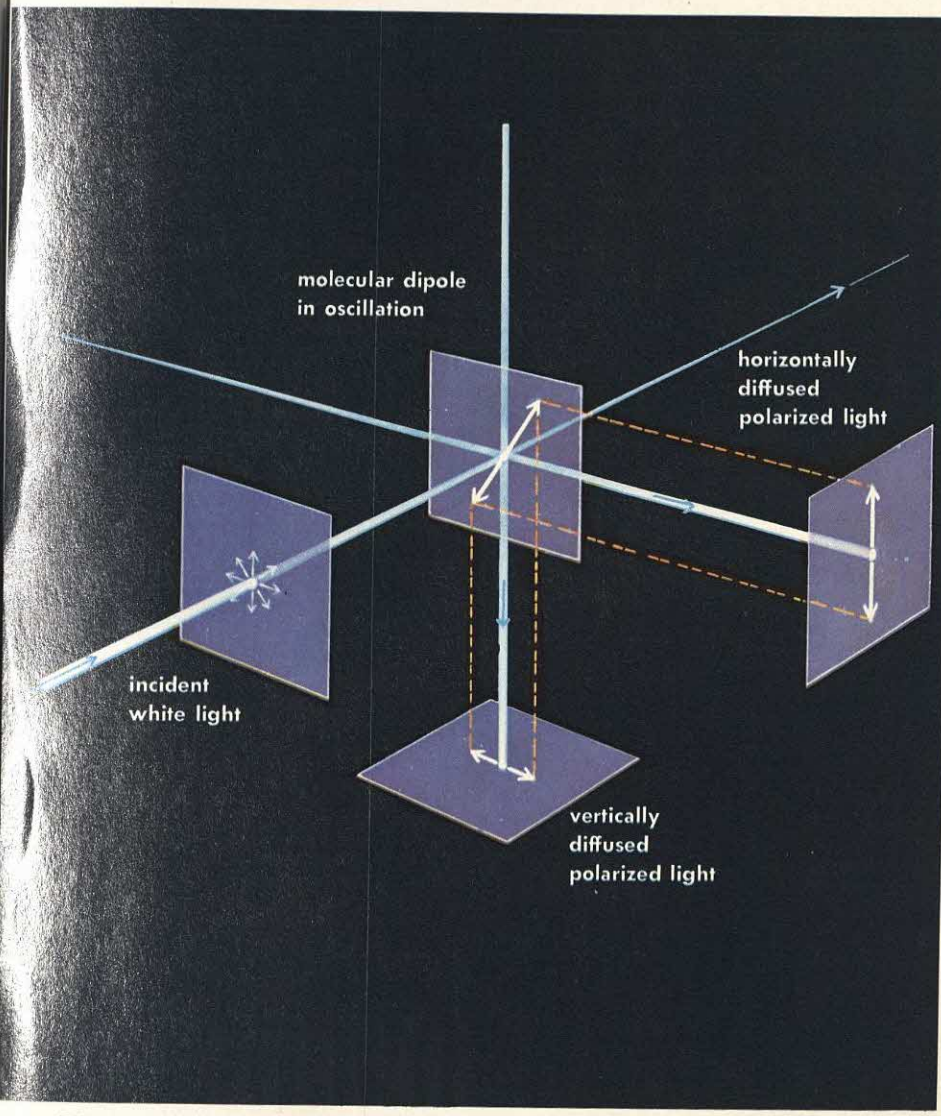
This article explains how light waves become diffused when they meet obstacles. An understanding of light diffusion permits the interpretation of many phenomena, some of which are discussed in this article. Moreover, an understanding of diffusion is basic to certain astrophysical measurements and leads to many practical applications.

The diffusion of sunlight accounts for the blue appearance of the sky and the reddish appearance of the sun at sunset. The light from most distant stars appears red because it is diffused by fine dust particles distributed throughout space. On a radar screen, hail may be detected falling from a storm cloud, and this observation is made possible by the phenomenon of light diffusion. Finally, this phenomenon permits microscopic observation of particles that are too small to be seen otherwise.

The study of light diffusion is quite complex and calls for the use of difficult mathematical theories. This article, however, is limited to a study of the fundamental principles of diffusion. Its goal is to provide the reader with an intuitive understanding of phenomena for which physicists have constructed complicated scientific explanations.

The goal may best be achieved through the use of an analogy—that suggested in the first paragraph of this article—comparing the behavior of light waves striking a grain of dust or a drop of water and the behavior of ocean waves breaking against a rock or pole or some other obstacle. In general, caution must be exercised in comparing phenomena of electricity or light with mechanical phenomena. However, in this case, no errors will be committed if the analogy is used to interpret particular groups of phenomena rather than to explain the nature of light.

2



DIFFUSION BY MOLECULES—Light-diffusion phenomena can be analyzed in different ways according to the dimensions of the particles causing them. The smallest diffusing particles are the molecules of a gas. The English physicist Lord Rayleigh was the first to study gas molecules from this point of view. He approached the problem by observing that the positive and negative charges contained in a molecule become displaced when a light wave passes and that they begin to vibrate in unison with the wave; this vibration results in the emission of light. Although the light that produces the vibration is propagated in one particular direction, however, the light emitted by the molecule is radiated more or less in all directions. Lord Rayleigh's studies showed that the extent to which a molecule diffuses radiation depends on the wavelengths of the light rays striking it; more precisely, light having short wavelengths is diffused more extensively than light having long wavelengths.

Lord Rayleigh also established the exact laws that govern the diffusion of various radiations and found that diffusion is inversely proportional to the fourth power of the wavelength. For example, red light has a wavelength of about 8,000 Å, and violet light has a wavelength of about 4,000 Å; the ratio between these two wavelengths is 2. The violet light is diffused more than the red and its diffusion is proportionally greater, to the fourth power of this ratio—that is to say, the violet light is diffused 2^4 or 16 times more than the red light. To put this another way, the red light is diffused only 1/16 as much as the violet light.

3

WHY THE SKY IS BLUE—The blue color of the sky (Illustration 3a) results from the diffusion or scattering of sunlight by the molecules comprising the Earth's atmosphere. Each of these molecules constitutes a diffusion center similar to the one shown schematically in Illustration 2; many molecules together diffuse the radiation in such a way that it is perceived as very intense light.

a

b

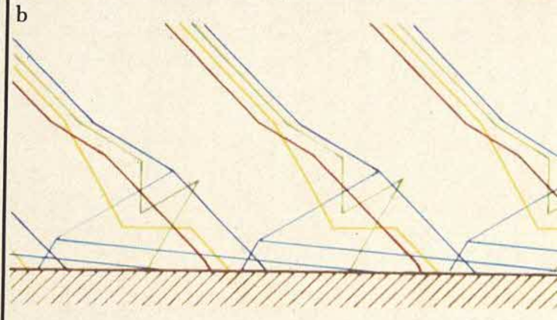
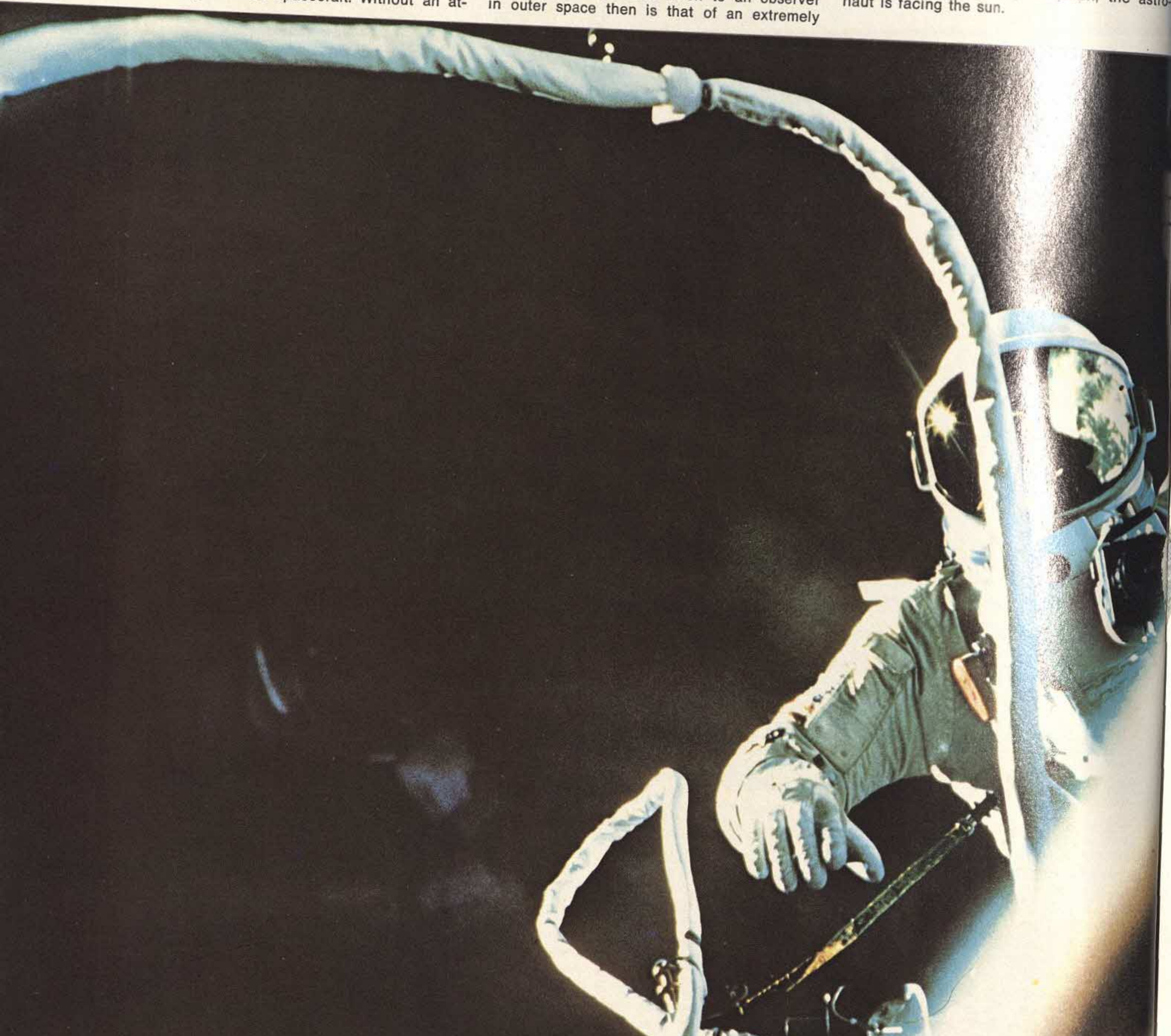


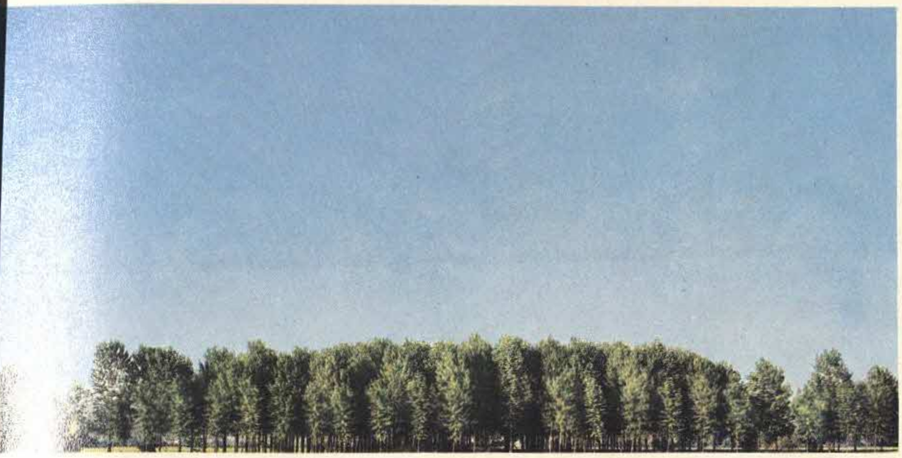
Illustration 3b is a schematic representation of the paths followed by light rays from the sun as they pass through the atmosphere. The solar rays strike molecules in their path and become partly diffused; therefore, they follow zigzag paths before they reach the eyes. Light from the sun is white light; that is, it consists of radiation of all the wavelengths comprising the visible portion of the electromagnetic spectrum. All these light waves become diffused when they strike molecules, but not all are diffused to the same extent. Rays of violet light are diffused the most, and rays of blue light are diffused only slightly less than violet rays. Rays of green light are diffused still less, and so forth, down to the rays of red light, which are diffused the least. Quantitatively, then, the violet and blue rays are scattered much more than the rays of the other colors. Violet light is not perceived well by the human eye; therefore, to the observer on Earth, blue is the predominant color of the sky.

4

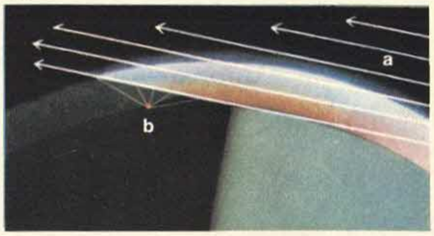
THE SKY AS SEEN IN THE ABSENCE OF AN ATMOSPHERE—This photograph shows the color of the sky as it appears at altitudes attained by manned spacecraft. Without an atmosphere the rays of sunlight cannot be reflected and diffused in all directions. The spectacle that presents itself to an observer in outer space then is that of an extremely bright sun entirely surrounded by a black (o colorless) sky. In this photograph, the astronaut is facing the sun.



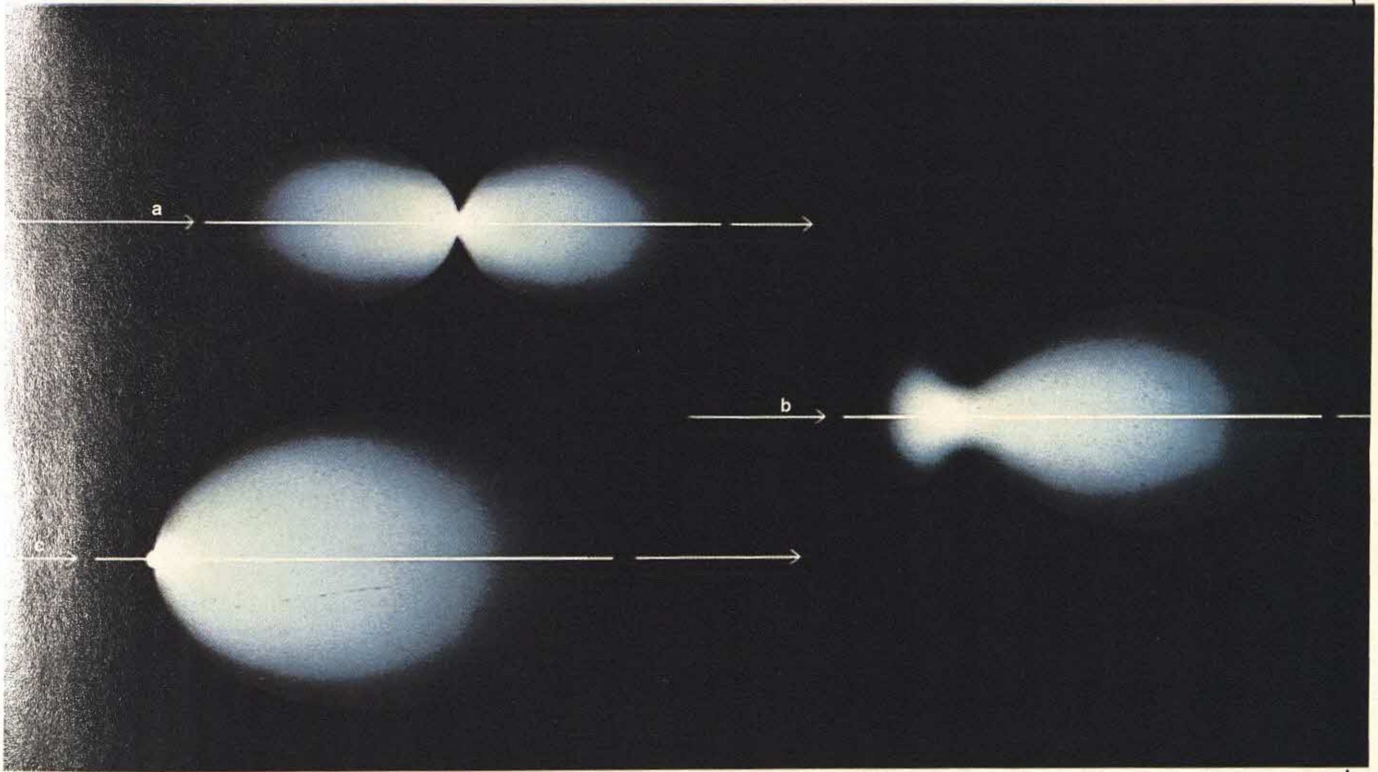
5a



5b



TWILIGHT—In accordance with Lord Rayleigh's laws, diffusion is also the cause of twilight (Illustration 5a). The diagram (Illustration 5b) shows why this is so. Rays of light from the sun **a** cannot reach the observer **b** directly because the sun is already below the horizon but they do reach his eyes because they strike the molecules in the atmosphere above him and scatter.



MIE'S THEORY OF LIGHT DIFFUSION—Some diffusion is caused by particles of matter whose dimensions are comparable to the wavelengths of the incident light. (These particles are much larger than the individual molecules that produce light diffusion in the atmosphere.) This type of diffusion can be caused, for example, by the particles of colloids in solution, by small, solid particles suspended in rivers, by the dust that fills interstellar space, or by the minute droplets of water that form a cloud or fogbank.

The manner in which the particles of a material diffuse the light striking them was studied in 1808 by the German chemist Gustav Mie. These complex studies are of fundamental importance because they serve as the basis for interpreting a great many phenomena. The complexity of Mie's studies notwithstanding, his results are quite easy to understand because they display a certain analogy with the behavior of waves in water. These illustrations demonstrate how the size of particles influences the diffusion of light.

Illustration 6a represents the diffusion of green light by gold particles having dimensions that are much smaller than the wavelength of this light—say 10 to 100 times smaller. In this case, about half the diffusion takes place in the direction in which the light is moving, and the other half is diffused backward; scattering is observed in these two principal directions.

Illustration 6b represents the diffusion of light by particles whose dimensions are comparable to the wavelength of the light—or even slightly greater. In this case, a large part of the light is diffused in the direction of the incident light ray.

Illustration 6c represents diffusion by even larger particles and shows that an even larger part of the light is diffused in the direction of the incident light ray. Mie's studies showed that, as the diameter of the particles increases, an increasingly large part of the light is diffused in the direction of the incident light ray; this is known as the Mie effect.

The phenomenon of diffusion is complicated by other factors besides the wavelength of the radiation and the dimensions of the intervening particles. It depends on the polarization of the light and on the conducting or insulating properties of the material of the particle. Many interesting phenomena can be predicted by means of this theory of diffusion. For example, the color that some colloidal particles assume in a solution can be predicted if their characteristics as electric conductors are known.

For quite a long period in 1950, the sun and the moon, when observed from many parts of the Northern Hemisphere, appeared to be light blue rather than their usual white. A detailed study of the light from these two bodies eventually enabled scientists to establish that the change of color was the result of a diffusion phenomenon caused by the smoke from a huge forest fire that had occurred some time previously in the province of Alberta, in Canada.

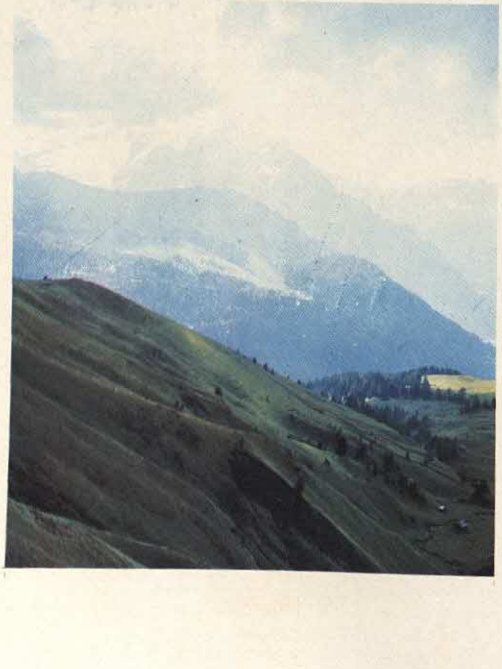
PHENOMENA OF LIGHT DIFFUSION

obstacles in the path of a light ray

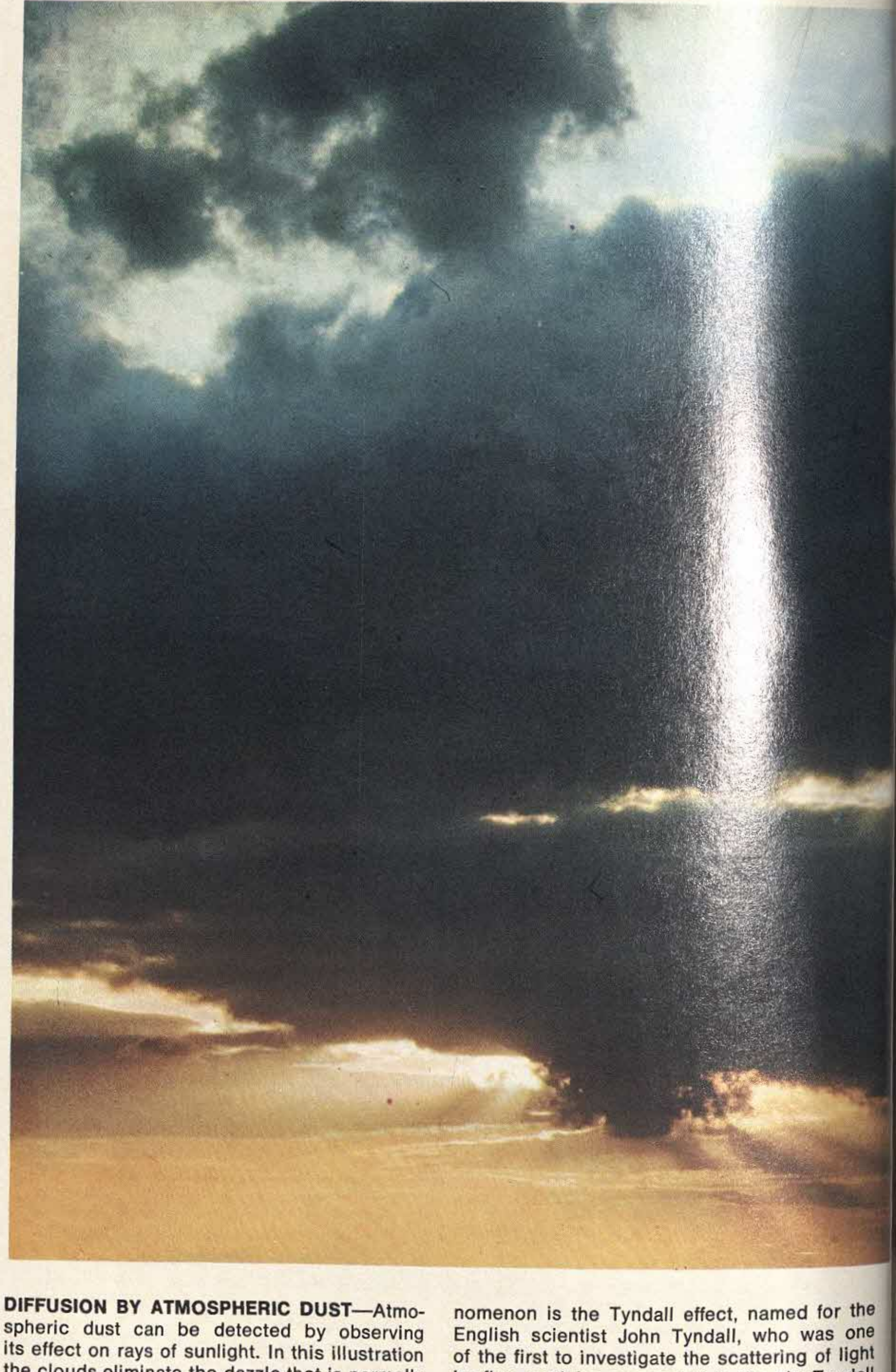
The diffusion of light rays when they strike small particles (so small that they are comparable to the wavelength of light) is a phenomenon familiar to everyone and readily observed in everyday life. It is interesting to be able to interpret this phenomenon and understand what causes all the effects that can be observed.

THE COUNTRYSIDE SEEN THROUGH A MIST—Mist prevents an observer from seeing distant objects with clarity. It may even hide them completely. This illustration shows how nearby objects may be sharply defined, with clearly distinguished colors, while objects in the background are much less distinct—their outlines can be perceived, but their colors merge into a uniform bluish gray. This is because of the diffusion of the light rays by minute particles of dust and water in the atmosphere; the water particles are similar to those of fog, but smaller. Rays coming from the distant objects have to pass through more layers of atmosphere. As a result, they encounter a larger number of particles that disturb their propagation and cause them to diffuse before they reach the observer. The effect produced by mist is thus much the same as that produced by ground glass: all the light manages to pass through it, but in so doing, it undergoes a very substantial dispersion. The distant countryside appears as a uniform background image—a direct result of the progressive mixing of the light rays reflected by distant objects.

1



2



DIFFUSION BY ATMOSPHERIC DUST—Atmospheric dust can be detected by observing its effect on rays of sunlight. In this illustration the clouds eliminate the dazzle that is normally caused by the sun. Here, the rays of sunlight that break through the cloud illuminate particles of atmospheric dust, causing part of the light to become diffused. This diffusion phenomenon is the Tyndall effect, named for the English scientist John Tyndall, who was one of the first to investigate the scattering of light by fine particles suspended in the air. Tyndall made his initial observations with a glass tank filled with a colloid; the illustration shows the same effect observed in the atmosphere.

A RED SUNSET—The magnificent yellow, orange, and red colors of the sky at sunset are explained by the theory of Rayleigh scattering (that scattering is inversely proportional to the fourth power of the wavelength, as pointed out by the English mathematician Lord Rayleigh). Thus, the violet and blue rays of the sun's white light are strongly diffused, while the red, orange, and yellow rays pass practically undisturbed, or at most, are only slightly diffused. At sunset, the light rays travel long distances in the lower layers of the atmosphere, layers that are richest in particles of dust or water.

Particles diffuse blue and violet rays, and the light, having little of these two colors, appears yellow or orange. The intensity of red coloring depends on how much dust is encountered along the light's path. At sea level, as shown in Illustration 3a, this phenomenon is very intense. Certain kinds of dust can increase the intensity of red. A sunset photographed at the edge of the Sahara Desert (Illustration 3b) appears particularly red due to sand particles suspended in the air and carried along by desert winds.

YELLOW HEADLIGHTS FOR USE IN FOG—Yellow headlights are more practicable than white ones for driving in the fog. The driver, regardless of whether he is driving at night or through fog, seeks to illuminate the sides of the road and obstacles in his way. In perfectly clear air, headlights of any color, especially white ones, illuminate all obstacles clearly and in detail.

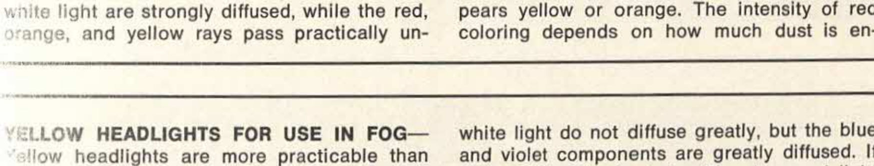
Fog diffuses the beam of light from the headlights, creating a kind of luminous barrier in front of the car. The road is illuminated partly by the light coming directly from the headlights and partly by light rays diffused by the fog particles. The light reflected by the objects, in turn, has to pass through the fog bank before it can reach the driver's eyes: this weakens visibility still further.

The poor visibility a driver experiences in fog is due to light diffusion, which decreases as the wavelength of light increases. The yellow, red, and green components of a beam of white light do not diffuse greatly, but the blue and violet components are greatly diffused. If the road is illuminated with a beam of light that has passed through a yellow filter, the components most readily diffused are eliminated (blue and violet). Clear visibility is the result because the barrier effect in front of the driver's eyes is reduced to a minimum. Obviously, only the yellow lights must be used; in fact, if white lights are kept on at the same time, poor visibility is not eliminated.

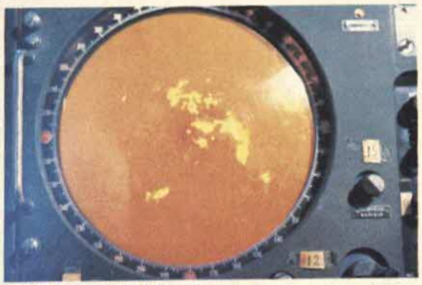
Why aren't orange or red lights used, since they have even greater wavelengths than yellow? There is less diffusion with these colors; however, the eye is less sensitive to red than yellow. Also, there is a danger of confusing red headlights with other lights used as traffic signals.

Today, sodium vapor lamps (which give off a yellow light) are commonly used to illuminate junctions that are often subject to fog.

4

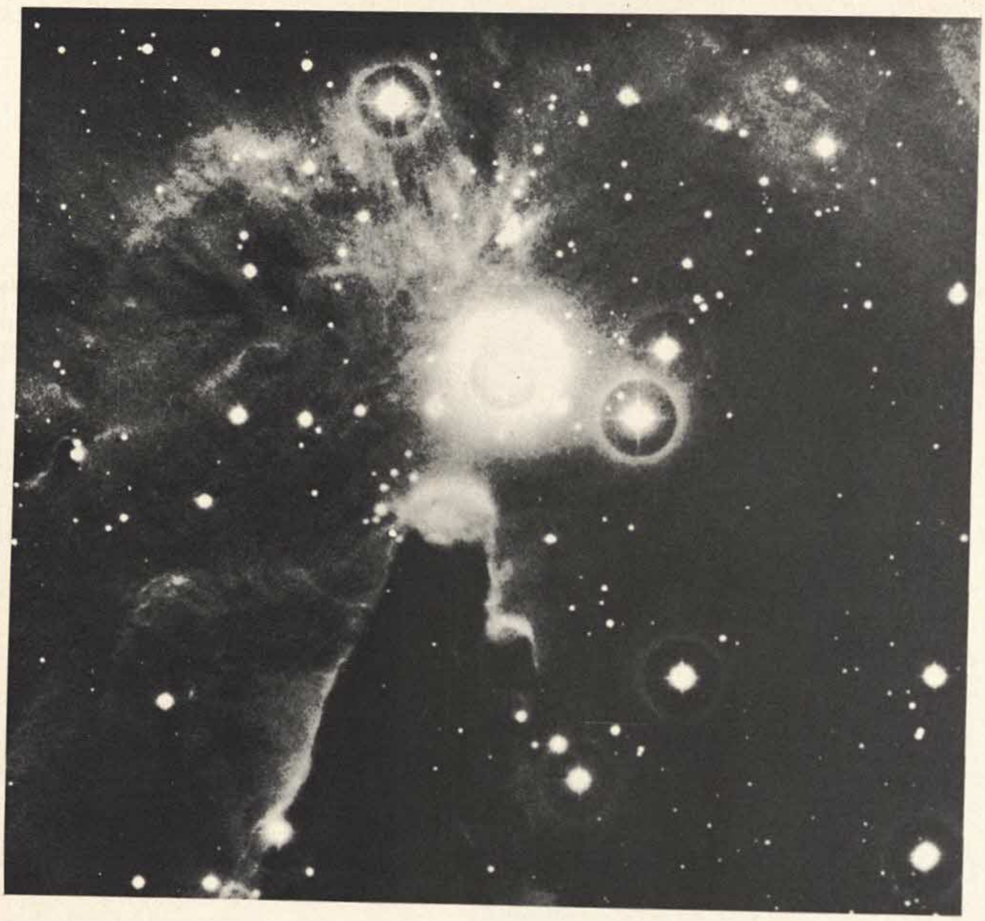


5



WEATHER RADAR—A radar set emits a beam of microwaves. When the microwaves strike a metallic obstacle or—more precisely—any conductor of electricity, they are reflected back and picked up by the same apparatus that previously emitted them. Microwaves have the advantage over ordinary light in that they can pass through clouds or banks of mist or fog. Even a microwave beam becomes somewhat attenuated, however, when it passes through thick layers of atmosphere in clouds or fog. This is because water droplets cause a certain amount of diffusion of the microwaves. It is critical to know the effect of these diffusing elements in order to interpret correctly the return signals received by the radar instrument. Moreover, if wavelengths of less than 3 cm are used, certain kinds of meteorological formations can alter the propagation of the waves. Extremely dense clouds, particularly clouds in which hail or snow is forming, diffuse a large percentage of the incident microwaves. This fact is the basis for investigating weather conditions that are dangerous for air navigation. A weather radar instrument is installed in the nose of aircraft to give advance warning of snow or ice formations in front of the aircraft.

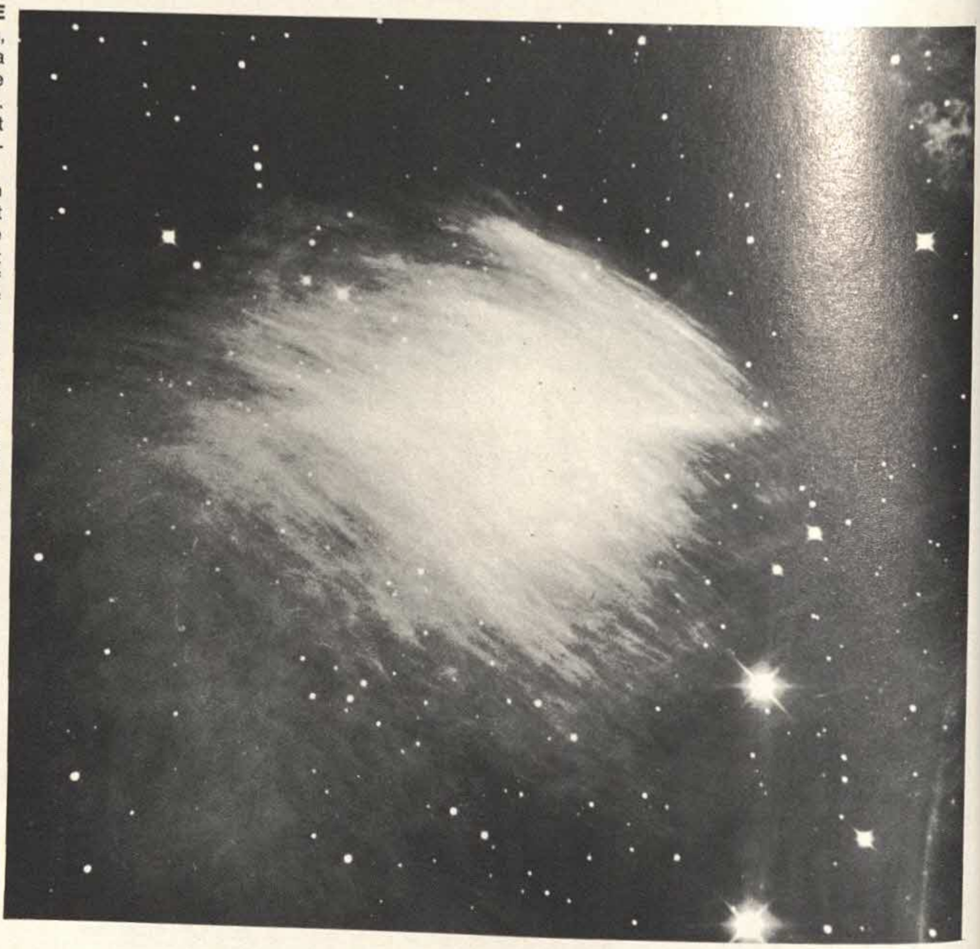
6



INTERSTELLAR ABSORPTION—In Illustration 6 view, the heavens are dotted with innumerable stars; nevertheless, there are strange gaps—completely dark areas that seem to be without any stars. It is assumed that the huge quantity of stars are distributed more or less uniformly. However, there are spaces between the observer and some stars that contain concentrations of fine dust that absorbs the light of the stars, in the manner of localized clouds.

Stars seen through these dust clouds seem redder than they really are; this phenomenon is analogous to that by which the sun appears to be red at sunset. Light is diffused as it passes through the dust clouds and, as shown here, blue and violet wavelengths are diffused the most, while red wavelengths manage to pass through the cloud.

7



DIFFUSION OF THE LIGHT OF THE STARS—Merope, a star in the Pleiades, is surrounded by dust that assumes a highly luminous appearance due to the diffusion of the light passing through it. Analysis of the light diffused by the dust will provide information about the nature of this dust.

First, the difference in color between the light emitted by the star and that diffused through the dust indicates the dimensions of the particles of the dust (which in this case are of the order of magnitude of a few ten thousandths of a millimeter). Furthermore, light diffused by these particles is often polarized, that is, the electric and magnetic vectors vibrate in a particular plane. This phenomenon indicates that the dust consists of elongated grains of magnetic material (like that of meteorites, iron and nickel) and that there are magnetic fields in space that align these dust grains.

LIGHT ABSORPTION | sunlight and water

A mysterious dark world exists in the ocean depths. A deep-sea diver can distinguish little in the nocturnal darkness. Yet, if the same diver swims near the surface of the ocean, he is able to see clearly. Here the sunlight penetrates the water and illuminates his surroundings. If the water is clear at all depths, then why does sunlight not penetrate the water to the ocean floor?

The answer is that some of the sun's light is absorbed by the water. An optical law governs the amount by which the sun's light is diminished the deeper it reaches into the sea. This law can easily be demonstrated when certain optical properties of water are known, and it can be appreciated without going deeply into the causes of the absorption of light by the water, or by the impurities contained in water.

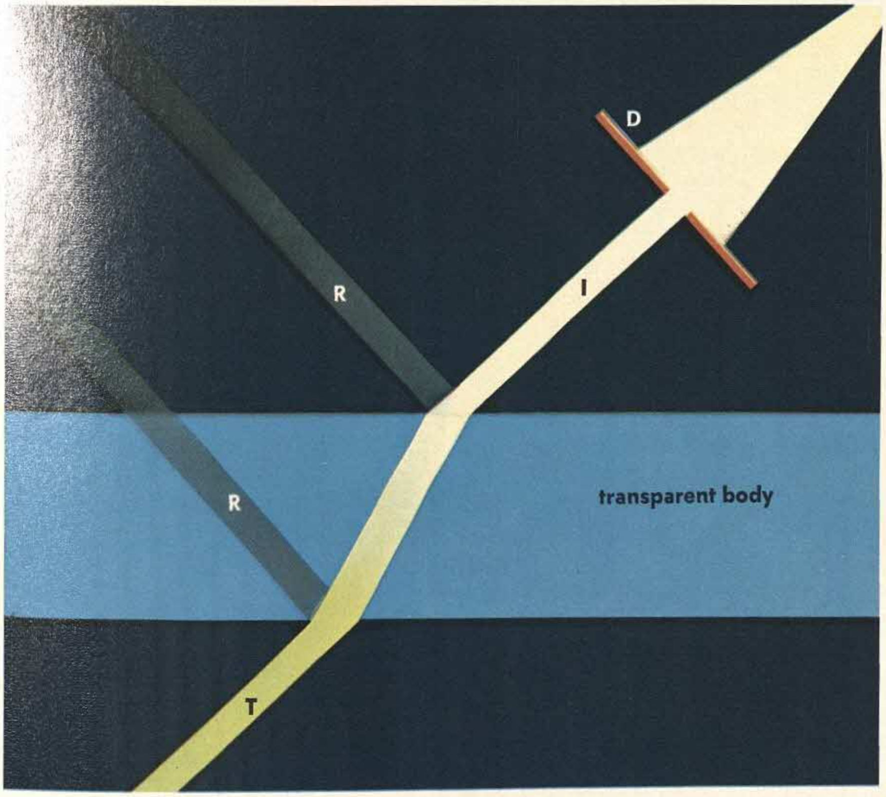
At first glance, research into such a subject might appear to be mere idle curiosity. Why would anyone be concerned over the fact that the ocean's floor is dark? The same law that governs the diminishing intensity of light as it penetrates a transparent body has many applications in physics and technology. All

LIGHT AND TRANSPARENT BODIES—No bodies are perfectly transparent. All bodies absorb at least some of the light that passes through them. On the basis of their thickness and the relative amount of light they absorb, bodies are classified as transparent, translucent, or opaque.

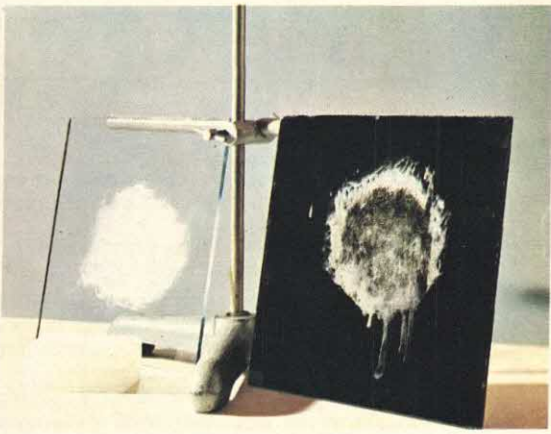
This illustration shows what happens to a beam of light when it strikes a transparent body (the degree of transparency does not matter). The beam, coming from a lamp and collimated by the aperture **D,** enters the transparent body—in this instance, colored water. As it enters, some of the light in the beam is reflected by the upper surface of the water, while the remainder penetrates it. A part of the light of this penetrating beam is absorbed. The layer of water here is not very deep, and the portion of the light that is not absorbed traverses the water until it strikes the opposite side. Before emerging from the lower surface of the water, the light beam is divided again. Some of it is reflected back into the water, while the remainder continues on its way into the surrounding air. In general, a beam of light **I** that traverses a transparent body is divided into three parts.

One of the parts **R** is reflected, another of the parts **T** is transmitted, and the third part **A** is absorbed and becomes invisible. The intensity of radiation **T** that emerges from the absorbing layer, plus the intensity **R** reflected by the interfaces, plus the amount of light absorbed **A,** equals the intensity of the incident radiation **I**; or symbolically $\mathbf{I} = \mathbf{T} + \mathbf{R} + \mathbf{A}$. This important law merely states that the luminous energy of the incident beam is divided into three parts, and that the sum of these parts is equal to the energy of the initial incident radiation.

1



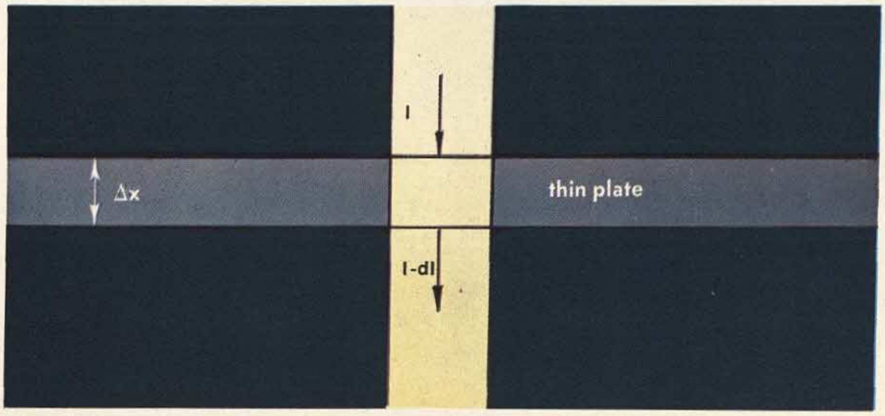
2



ABSORBED RADIATION—Many branches of physics and chemistry are concerned with the radiation absorbed by transparent substances. Certain substances, for example, undergo chemical reactions through the effect of light. The point to which the reaction has proceeded and the amount of the substance that has reacted are directly proportional to the amount of light that has been absorbed and not to the amount that has passed through the substance. Even simpler effects, such as the heating of a substance by radiation, depend on the amount of radiation absorbed.

The illustration shows two sheets of glass of the same size. One is transparent and the other is darkened. Both are exposed to the sun for the same length of time. Each receives the same amount of luminous energy on its surface, yet the dark sheet of glass will become hotter than the more transparent one. This difference in temperature can be emphasized by smearing a little paraffin diluted with paraffin oil on each sheet. The paraffin liquefies easily on the dark sheet of glass but remains solid on the clear one, showing that the amount of light absorbed by the transparent glass is less than the amount of light absorbed by the opaque glass.

3



ABSORPTION BY A THIN PLATE—This illustration shows a thin plate struck by a luminous beam of known intensity. Neglecting the amount of light lost through reflection at the surface as the beam enters the plate, the amount of light absorbed has the following characteristics:

1. It is proportional to the amount of light striking the plate.
2. It is proportional to the thickness of the plate (a thin plate will absorb less light than a thick one).
3. It also depends on the material of which the plate is made.

These relationships can be represented by a single formula: $-\Delta I = c \times I\Delta x$, in which I represents the intensity of the light striking the absorbing plate, and $-\Delta I$ the decrease in intensity due to the amount of light absorbed by the thin plate, the thickness of which is symbolized by Δx. c is a constant for any given material of which the light-absorbing plate may be composed; it is known as the coefficient of absorption and characterizes the capacity of the material for absorbing light. If the thickness of the absorbing plate, the intensity of the incident light, and the optical characteristics of the material—that is, its coefficient of absorption—are known, it is possible to predict the amount of absorption that will take place. As it stands, this formula is valid only for extremely thin layers of material. A more complex formula is required when the thickness of the layer is appreciable.

chemical analyses performed with spectrophotometers and colorimeters depend on this law of absorption; the law also applies to the absorption of other radiations, such as x-rays and gamma rays. These radiations, like light, are electromagnetic waves, although they have much shorter wavelengths. An engineer calculating the thickness of lead or concrete required to shield workers from radiation in a nuclear plant makes use of the law of light absorption.

TRANSMISSION AND OPTICAL DENSITY

The intensity of the light transmitted by an absorbing medium is measured by instruments known as photometers. A photometer produces an electric current when exposed to the light from any source. The strength of this electric current is proportional to the intensity of the light. If a translucent body is placed between the source of light and the instrument, the decreased current produced by the photometer will be in proportion to the amount of light getting through. The ratio of the amount of light transmitted to the amount of light that strikes the surface of the absorbing body is known as the transmission of the body. Because the amount of light emerging from the body is always less than the amount of light striking it, the transmission will always be a number smaller than 1. It can also be expressed as a percentage, by multiplying the number by 100. The lenses of a pair of dark glasses, for example, allow only a quarter of the light that strikes them to pass through. Their transmission is, therefore, 0.25 or 25 percent.

Another way of expressing a body's transmission is through its optical density, which is defined as the logarithm of the inverse of its transmission. For example, consider a sheet of opaque glass that transmits only 10 percent of the light that strikes it. This glass has a transmission of 0.1. The optical density, or opacity, of this sheet of glass is therefore 1. Another sheet of glass, transmitting only 1 percent —that is, reducing the light that strikes it one hundred times—has an optical density of 2. The glass that lets only a thousandth of the light through has an optical density of 3. These numbers—1, 2, and 3— are the exponents of the number 10 or the power to which the number ten must

ABSORPTION BY THICKER BODIES—The formula discussed under Illustration 3 cannot be used directly to compute the amount of light absorbed, for example, by the deep sea as the light penetrates all the way to the sea floor. In this case, the absorbing layer of water is quite thick, and the light is absorbed in a gradual manner.

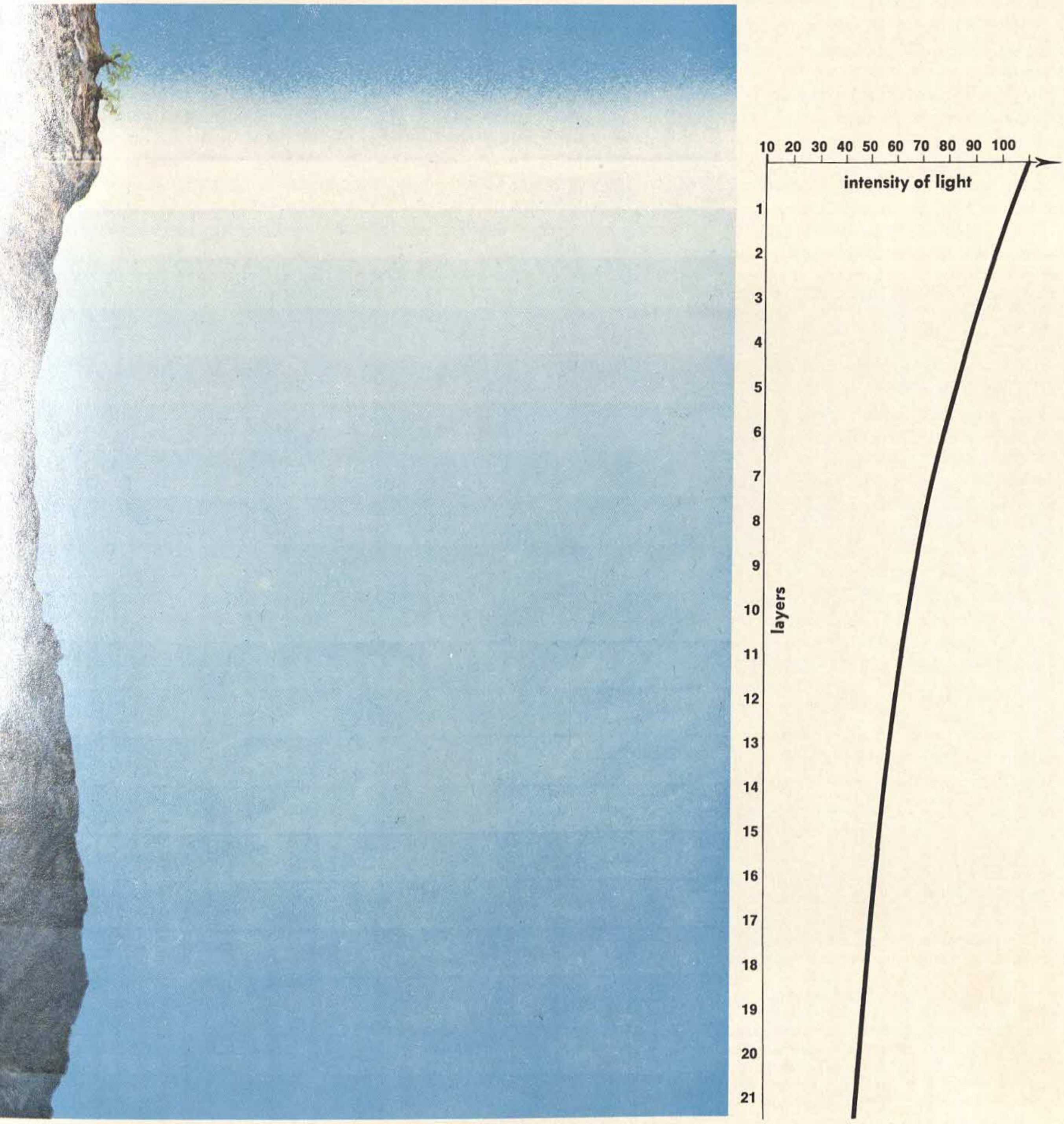
The problem can be solved by dividing the entire depth of water traversed by the light into a large number of thin layers adjacent to each other and equal in thickness. In moving through each one of these layers, the light is absorbed according to the formula given under the preceding illustration. Each layer absorbs a certain percentage of the total light that passes through it, and this percentage depends on the optical characteristics of the material; that is, its coefficient of absorption and the thickness of the layer. The layers of seawater are equal in thickness and all have the same coefficient of absorption. Thus, each diminishes the light striking it by the same amount, say 5 percent. It is, therefore, possible to draw a graph of the intensity of light found at different depths. If each layer absorbs 5 percent of the light, then 95 percent of the light passes through it. Therefore, only 95 percent of the amount of light striking the first layer will strike the second. The second layer, in turn, allows 95 percent of the light to pass through it. Thus, only about 90 percent of the light that strikes the surface of the sea will pass through the second layer. The light emerging from the third layer will be reduced to 95 percent of 90 percent, that is, to 85.7 percent, and so forth. After passing through 20 strata, the light will have been reduced to $0.95^{20} = 0.36$, or about 36 percent.

The illustration represents a body of water traversed by light. The accompanying graph shows that the light is reduced according to an exponential law. By definition, an exponential law is one that expresses a change of the same percentage on each successive occasion. Another example of an exponential law is the process of radioactive decay.

The formula that summarizes the individual effects of all of the many thin layers and correctly expresses the absorption of light regardless of the thickness of the absorbing layer is: $I_x = I_0 e^{-cx}$, where I_0 is the intensity of light striking the surface of the absorbing layer, and I_x the intensity of the emerging light. e is a mathematical constant equal to about 2.72, c is the absorption coefficient of the absorbing medium (in this case, seawater), and x is the total thickness of the layer (in this case, the depth of the sea). This formula can be used to calculate the absorption of light by any thickness of any medium, provided the thickness and the absorption coefficient are known.

be raised to obtain 10, 100, and 1,000.

The optical density of a body is a more useful concept than the transmission. When an opaque substance is dissolved in water, the optical density of the solution is proportional to the concentration of the substance dissolved. The optical density is used in chemical analyses to determine the amount of solute in a given solution. Chemical laboratories determine the concentration of many substances in solution by means of a spectrophotometer, an instrument that measures the optical density of the solution with reference to a beam of monochromatic light.



LIGHT INTERFERENCE | light does not always illuminate

A student who takes up the study of light is introduced first to geometrical optics. Armed with the knowledge that light travels in straight lines, he carries out experiments in which shadows are observed to have clear outlines and rays move in a straight line. He measures the angles at which light is reflected when it meets a mirror or the surface separating two bodies with different powers of refraction.

Geometrical optics is so named because it interprets the behavior of light rays as geometrical phenomena. The geometrical theory is adequate for solving a number of practical and theoretical problems; it is used in calculations concerning lenses and prisms, for example. Geometrical optics tells only part of the story, however; it neglects the extremely important fact that, by nature, light travels in waves.

The wave theory of light is known as physical optics to distinguish it from geometrical optics, and because it is concerned with the physical nature of light itself. The wave theory of light is important beyond the field of optics and is applied to radio waves, the design of antennas, of radio telescopes, and of a large number of extremely sophisticated optical instruments. Because physical optics provides the basis for an understanding of atoms and the motion of subatomic particles, it has led to the fundamental concepts of quantum mechanics —the theory that explains the behavior of atomic systems that do not obey the laws of classical mechanics.

SOME COMPARISONS WITH MECHANICS

The great developments in mathematics at the end of the eighteenth and the beginning of the nineteenth centuries led to the development of extremely accurate theories for the prediction of mechanical phenomena. The movements of the stars and planets were worked out. The motions of the waves on the sea's surface were studied. Theories to predict the motion of a pendulum attached to another pendulum were devised, as well as theories concerning the behavior of two colliding billiard balls.

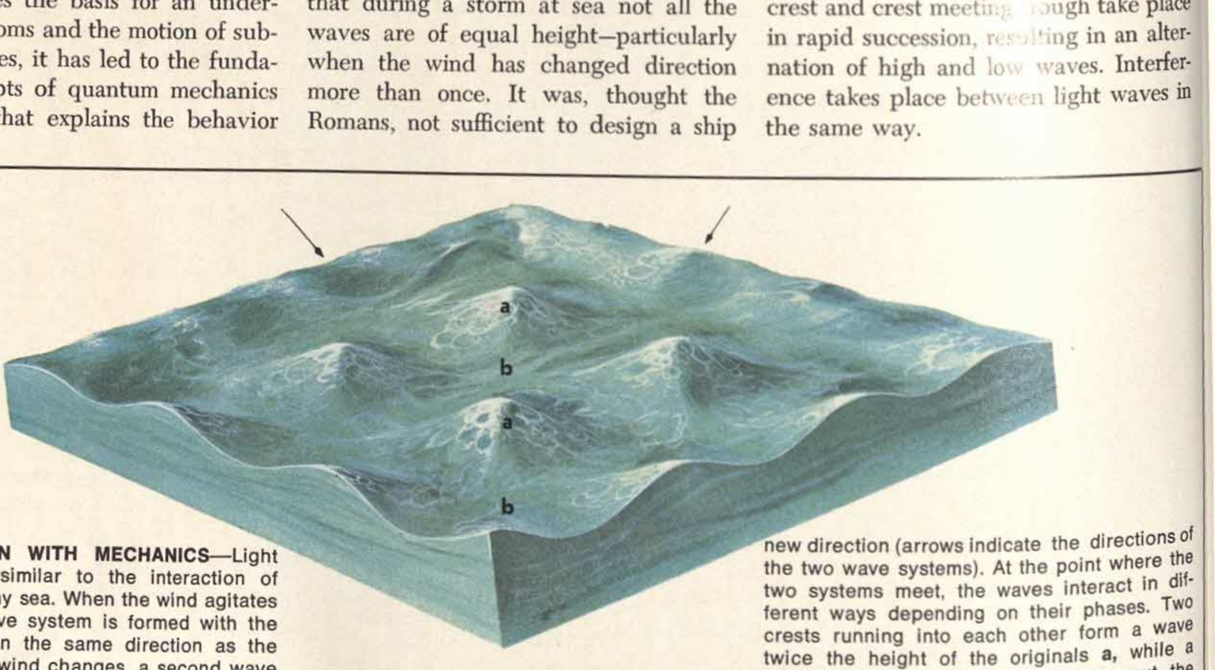
This great faith in classical mechanics stimulated scientists to work out the mathematical behavior patterns of many physical phenomena. However, classical theory proved to have a number of limitations when applied to certain phenomena. Still, some of the methods of classical mechanics are useful in helping to explain the movement of light waves and in understanding the nature of light interference.

The ancient Romans, who were fairly accomplished navigators, had observed that during a storm at sea not all the waves are of equal height—particularly when the wind has changed direction more than once. It was, thought the Romans, not sufficient to design a ship that would stand up to the average wave height because the ship would encounter much bigger waves from time to time. The Romans believed that every tenth wave was much higher than the others; they called this wave the decumanal wave. It is not true, of course, that one wave in ten is higher than the others, but the Romans were correct in noting that after a change of wind, all waves are not of equal height. Periodically, one wave is very much higher than the rest. This phenomenon can be explained by the effect known as interference.

When the wind agitates the sea, all the waves move in the direction in which the wind is blowing. As soon as the wind changes direction, a second wave system is created, running in a new direction. When the two wave systems meet, the waves will either combine or tend to cancel each other out depending on the phase of each wave at the moment of contact. The phase of the wave is the point to which it has developed: crest, trough, node (average level of sea before disturbance began), or any intermediate point. Two crests coming together create a wave twice the height of the original waves. Contact between a crest and a trough flattens the wave out.

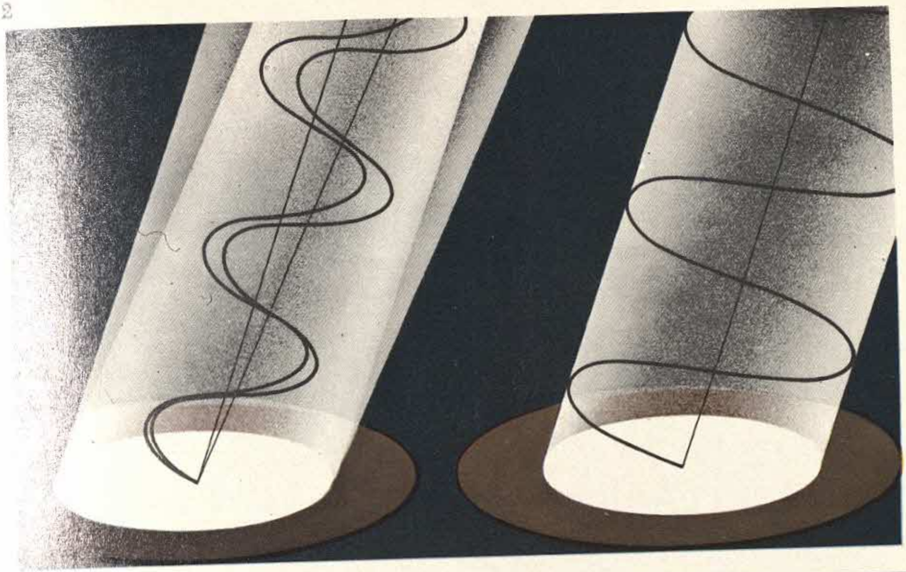
At sea, the situation of crest meeting crest and crest meeting trough take place in rapid succession, resulting in an alternation of high and low waves. Interference takes place between light waves in the same way.

1



A COMPARISON WITH MECHANICS—Light interference is similar to the interaction of waves in a stormy sea. When the wind agitates the water, a wave system is formed with the waves running in the same direction as the wind. When the wind changes, a second wave system is formed with the waves running in a new direction (arrows indicate the directions of the two wave systems). At the point where the two systems meet, the waves interact in different ways depending on their phases. Two crests running into each other form a wave twice the height of the originals **a**, while a crest running into a trough flattens out the wave **b**.

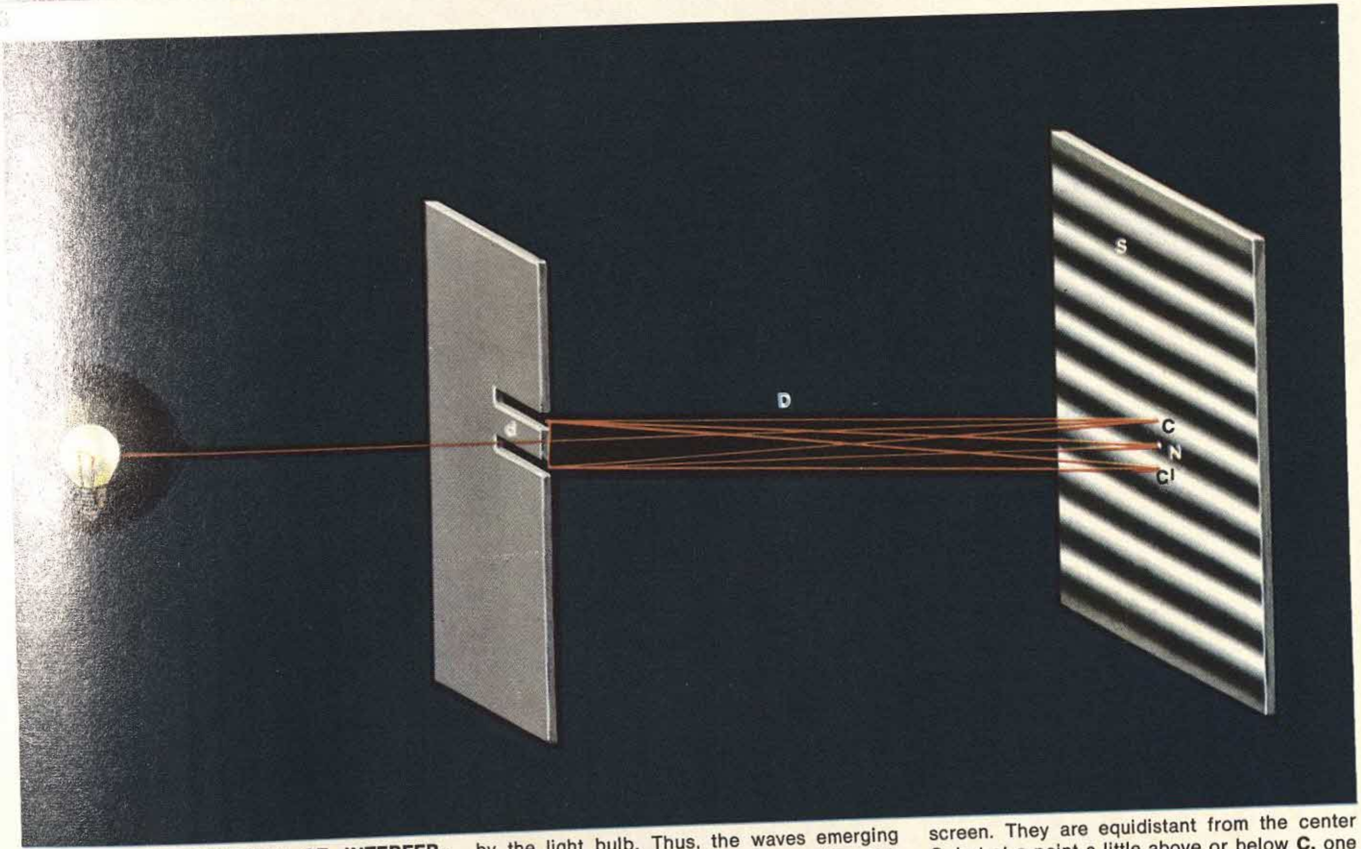
2



THE INTENSITY OF LIGHT—The intensity of light can be interpreted according to the wave theory of light. Light intensity depends on the strength of the light vibrations and is proportional to their square.

A light ray is an electromagnetic vibration that is propagated in space like the sinusoidal wave shown in the illustration. At the left, the central part of a screen is illuminated simultaneously by two light sources. The two sinusoidal vibrations, in phase as they meet the screen, are added together and produce a stronger light than one light source could produce. Two vibrations are said to be in phase when they are moving so that crest corresponds to crest and trough to trough. At the right, two waves have become united to form a single hypothetical vibration having an amplitude double that of the original waves. This phenomenon is known as constructive interference: constructive because the sum of the two produces a more intense light, and interference because the two waves have united to form a single wave.

3



PROVING THE EXISTENCE OF INTERFERENCE—Illustration 2 showed how the sum of two light waves can be interpreted on the basis of the wave theory of light. Similarly, two light waves diminish in intensity if they meet out of phase. This experiment demonstrates the phenomenon of interference. It is carried out with two narrow, parallel openings illuminated by a single source (an electric light bulb) a set distance away. Each opening, therefore, becomes a light source; each emits light waves in phase with those emitted by the light bulb. Thus, the waves emerging from each opening are in phase with each other and can be represented by sinusoidal curves with their crests aligned.

The screen **S,** placed at a given distance **D,** is illuminated not uniformly but in bands. The illustration suggests a simple explanation for this phenomenon: the two openings, separated by a distance **d,** are struck by a single wave front and thus emit radiations that are in phase. However, the two openings are not at the same distance from every point of the screen. They are equidistant from the center **C,** but at a point a little above or below **C,** one is more distant than the other. At the position **N,** the difference between their distances corresponds to exactly half a light wavelength. At this point, the waves meet crest to trough and destructive interference results, producing darkness. At **C'** the difference in their distance has been increased to a whole wavelength; the light waves are again in phase and the result is a bright band of light.

4

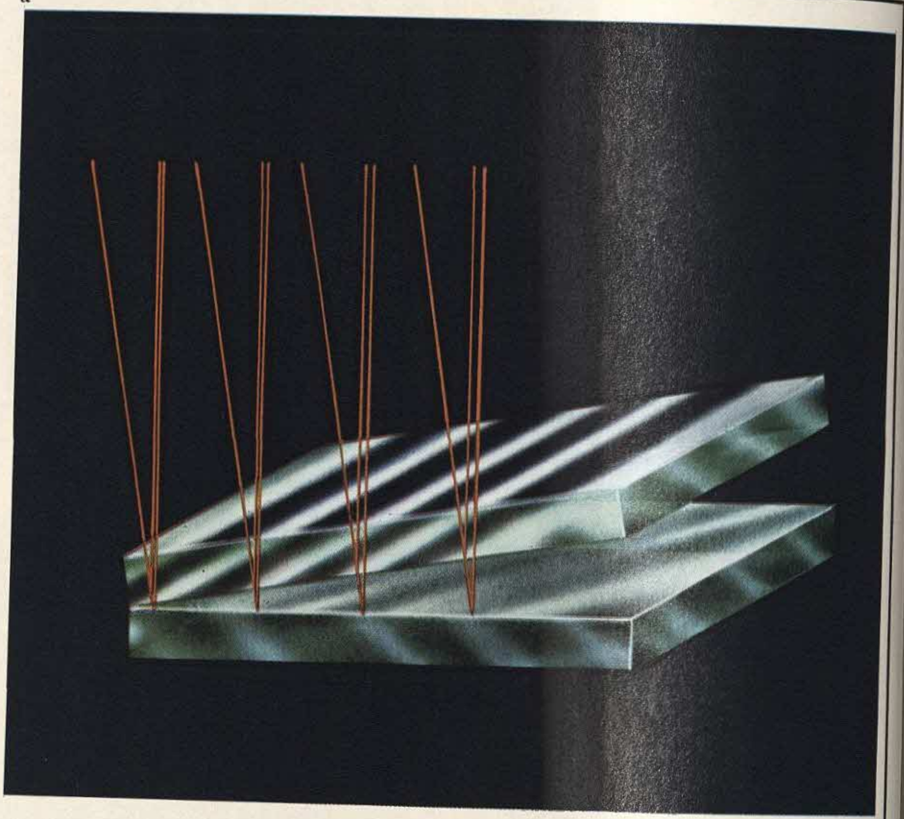
AN IMPORTANT APPLICATION OF INTERFERENCE—Light interference occurs in many different ways; it can occur every time two light waves have vibrations of the same wavelength but are out of phase by a whole number of half wavelengths.

One way to obtain interference of light waves is shown in Illustration 4a. Two plates of glass have been placed at a very small angle to each other. If a beam of light is directed from above and only the two facing surfaces of the glass plates are considered, the light is reflected twice, once by each plate. The light reflected by the surface of the upper plate returns sooner than that reflected by the surface of the lower, because the light reflected by the lower plate must span the whole distance separating the two plates twice (once as it arrives and again as it returns). The delay is variable, however, because the two plates are at an angle to each other and the distance between them is not the same at all points. The result is dark and light areas, consisting of bands parallel to the edges of the two plates—areas that are adjacent and in which constructive and destructive interference take place alternately.

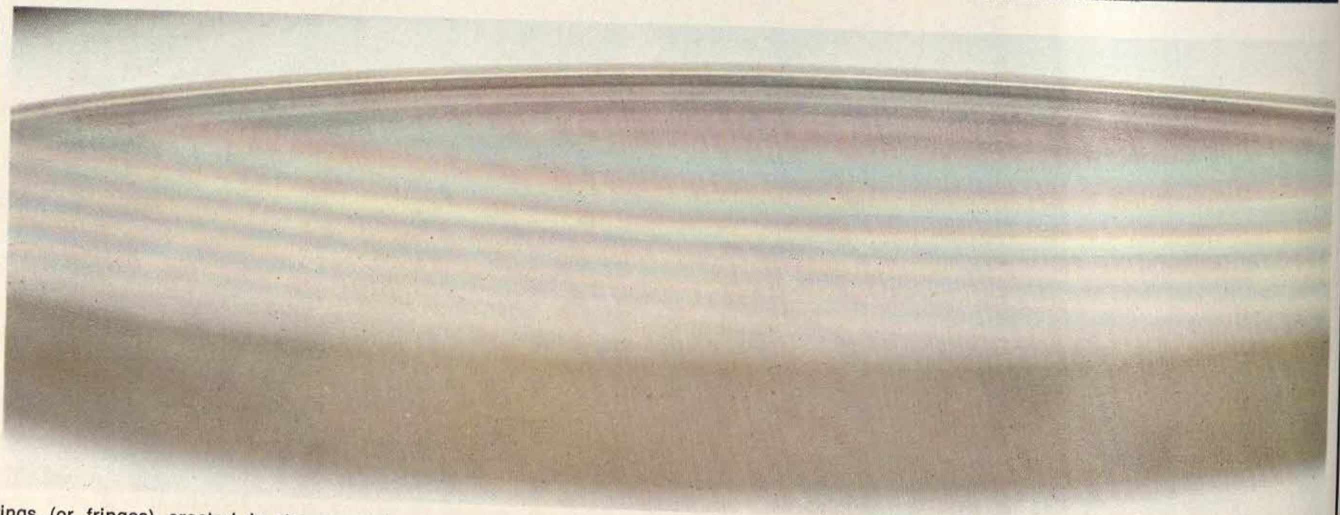
If spherical surfaces are placed in contact, the same effect results, provided that the radii of their curvature are slightly different. In this case, the areas in which interference takes place are circular instead of rectilinear and are known as Newton's rings. The famous English scientist did not believe in the wave theory of light but gave a different interpretation of it. Illustration 4b shows Newton's rings (or fringes) created in the space between a flat and a concave surface when the surfaces are placed together. If the two surfaces were respectively convex and flat, they would produce perfectly circular rings.

To produce distortion of the rings, the two surfaces in contact need differ from the ideal form by only a small fraction of a light wavelength. In this way, it is possible to show deviations from the ideal flat or spherical form of as little as 0.00001 mm (about 0.0000004 in.). For example, the distortions of the two surfaces placed in contact to obtain interference in the illustration are no greater than 0.001 mm (about 0.00004 in.). This method of checking accuracy is used to test the spherical properties of lenses to be placed in optical instruments.

a



b



DIFFRACTION OF LIGHT

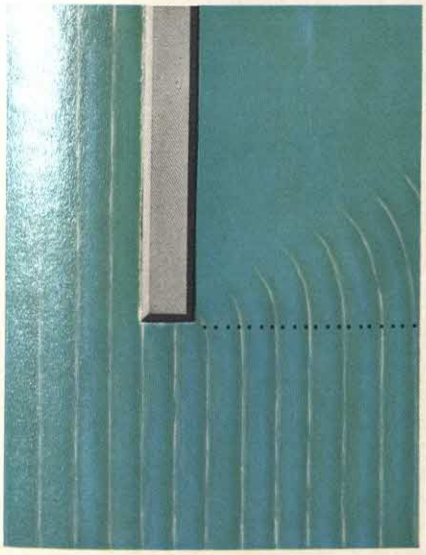
the wave theory

Light, it is commonly assumed, travels in a straight line—but in actual fact it sometimes behaves very differently. Light can, in fact, turn corners without encountering any obstruction in its path, and this phenomenon is caused by *diffraction*, which demonstrates that light travels in waves.

Light waves are different in nature from waves that are observed in a body of water. Light waves are electromagnetic; water waves are mechanical. However, waves on the surface of water can be used to illustrate the behavior of light waves and diffraction, according to Huygens' principle. Simply stated, Huygens' principle says that the front of each wave is a source of propagation for new waves.

All fields of physics make use of the principle of diffraction. The methods used, for example, in the x-ray study of the structure of matter are based on the principles of diffraction and interference; similar studies have been carried out on electrons and neutrons. If these particles are launched at the proper velocity through a crystalline material, they behave like waves rather than like particles of matter, undergoing diffraction, a phenomenon typically associated with waves. This particular example has opened the path to the study of the wave theory of matter and thus to atomic, or quantum, mechanics, an outgrowth of the concept that energy is released in discrete units.

THE DIFFRACTION OF A WATER WAVE—A dike, placed in the deep water of a basin, is hit head on by a wave. Where the dike stands in its way, the wave of course is blocked, but beyond the end of the dike it continues to move. Instead of proceeding in a straight line (along the dots), the wave fans out, reaching the part of the basin protected by the dike. This part of the basin might be expected to be calm. But the wave coming into it creates a disturbance that behaves exactly like a new source. From it, waves radiate in all directions. Beyond the form of surface waves in water, the term *waves* extends to cover any vibratory motion propagated through a medium. Diffraction of light happens in a similar way, although the phenomenon is more complicated.

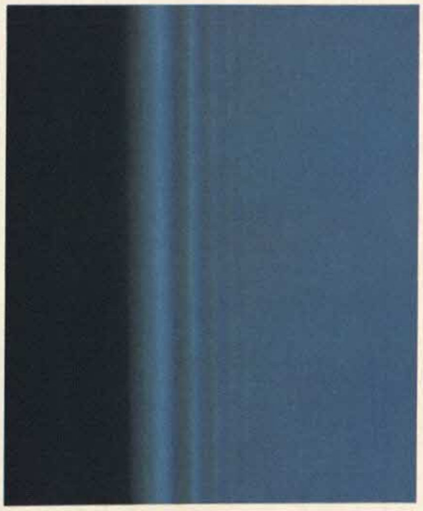


2

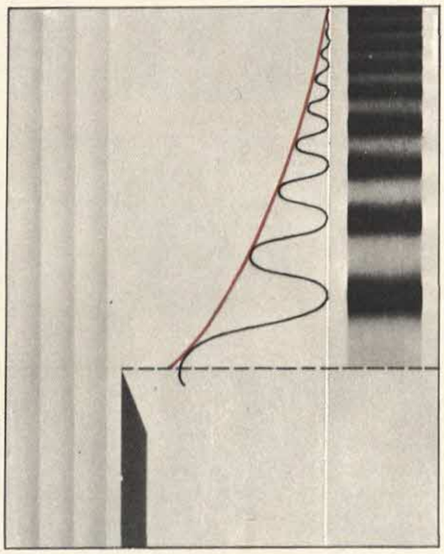
THE DIFFRACTION OF LIGHT—In Illustration 1, waves on the surface of water were breaking against a dike that was interrupted laterally. In Illustration 2a, light waves are obstructed by an opaque screen that is interrupted laterally. A point of light is placed at a distance of 1-2 m (about 3-6 ft) from a flat object with a well-defined outline, such as a knife or razor blade. On hitting the edge of the blade, the light wave will not proceed in a straight line but will fan out. If a screen, such as the photographic plate used here, is placed a short distance from the blade, the image of the diffracted light can be seen. The diffraction is more than the lateral fanning out of the light; a shadow results that is not clear but seems to shade off toward the darkness. Observe that before it diminishes, the light forms several more intense bands.

In Illustration 2b, the course of the light and darkness is examined perpendicularly to the length of the cutting edge of the blade and the flat side of the same blade. The illustration shows the strength of the light beginning from the illuminated part farthest from the cutting edge of the blade. Corresponding to the cut of the blade, the strength of the light (which was uniform before) drops suddenly, then rises again in the points of light beyond the blade to a value that is greater than that obtained if the light had arrived undisturbed. An explanation for this phenomenon is found in the application of Huygens' principle.

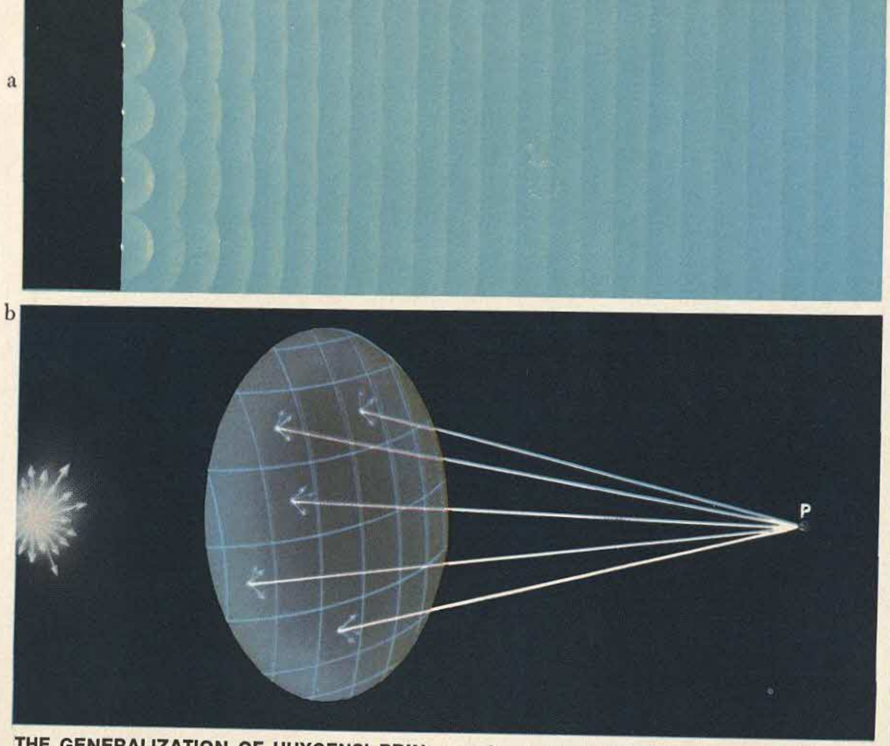
a



b



3



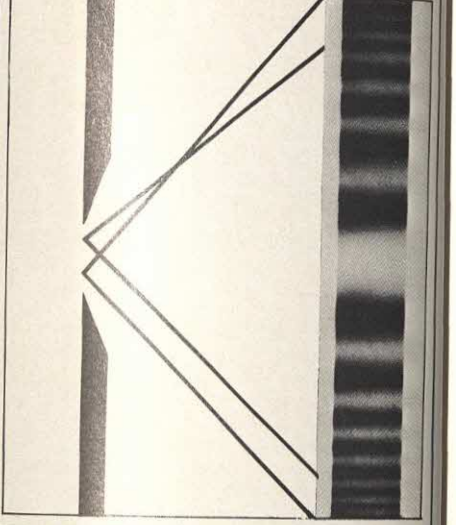
THE GENERALIZATION OF HUYGENS' PRINCIPLE—Huygens' principle affirms that a wave front is a source of disturbance in space. Each point on a wave front generates a spherical wave around itself. Imagine that the wave front is at rest at a certain instant; many small spherical waves are being generated from its surface (Illustration 3a). These new waves are stronger in the direction of propagation of the original light wave and weaker laterally to it. Added together, they form a new compact front. This, in turn, can be imagined as the source of many small waves.

To interpret the phenomenon of diffraction, it is convenient to state Huygens' principle in another way. Illustration 3b shows the arrangement, at a given instant, of the parallel wave fronts propagated by a distant light source. If they are allowed to propagate indefinitely, they will eventually reach the point **P**, which will then become illuminated.

Huygens' principle can be applied under a new form to show how the point will be illuminated. Divide the surface of the nearest wave front into a number of small areas: according to Huygens' principle, each would emit a light disturbance approaching the point **P** and striking it. Therefore, the illumination to which **P** is subjected depends on the light emitted by the various small areas into which the surface of the wave front has been divided. Those nearest to it will illuminate it most; the distant ones least.

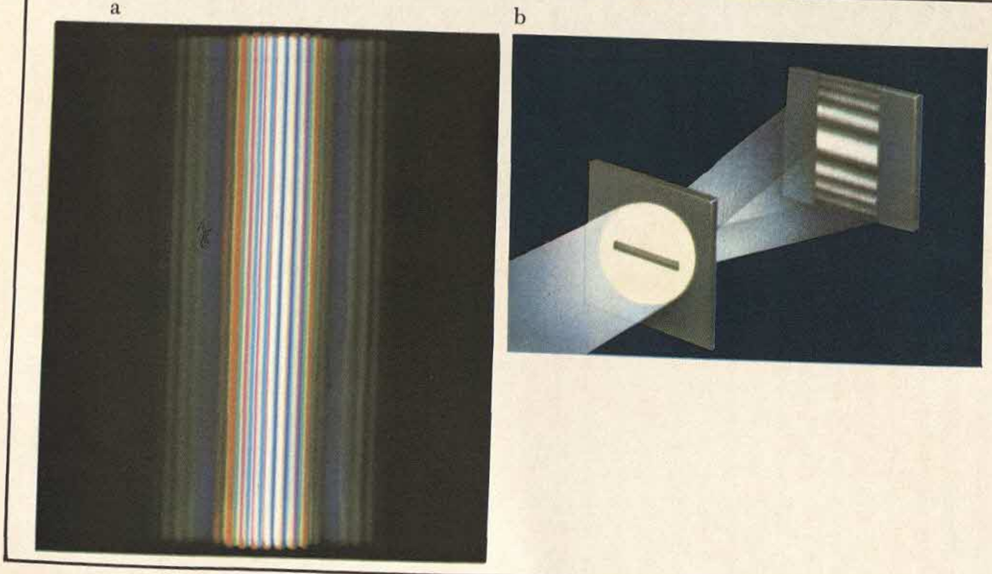
Note that there is a difference in distance to point **P** from the various parts of the wave front. Light emitted at the same time will reach **P** in phases. In other words, the waves reaching **P** will arrive in a disorderly manner. The crest of a wave from part of the front set will arrive simultaneously with the trough of a wave coming from another part. Each area sends out waves that arrive in different phases and at different strengths. The fundamental consequence is that if all the light from the different areas of the wave front arrive at **P**, this point will be illuminated because the light is added together, resulting in constructive interference. If, on the other hand, light from certain areas of the front is prevented from reaching **P**, the point may appear dark because part of the light reaching it may be annulled by interference.

5



THE EXPLANATION OF DIFFRACTION—The diffraction of light as shown in Illustration 4a can be explained by Illustration 5. A distant source emits a flat wave front (initially spherical but rendered virtually flat by distance) in the direction of the slit. When the light arrives, the slit acts as a screen, halting part of the wave front. Suppose that the slit is divided into at least two smaller horizontal slits, which, according to Huygens' principle, will become many new sources of phased light. When they reach the screen, the waves emitted by the various parts of the slit will create the phenomenon called interference. The light that goes through the slit does not simply illuminate a rectangular area, but forms a rectangle of light with shaded edges, accompanied laterally by other beams of shaded light. These beams are less intense because, according to Huygens' principle, light diffracted laterally has less strength.

4

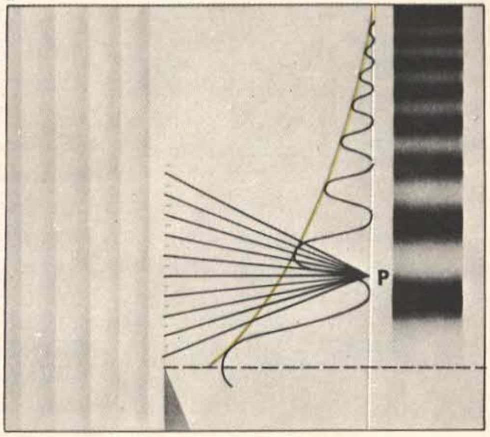


DIFFRACTION THROUGH A SLIT—Illustration 4a was obtained by illuminating a narrow slit from a source or point of light. The slit is not much larger than the wavelength of the light and is a long distance from the screen onto which the light is to be projected. The light that passes through the slit does not reach the screen in the form of a geometric rectangle, as the laws of geometrical optics would require.

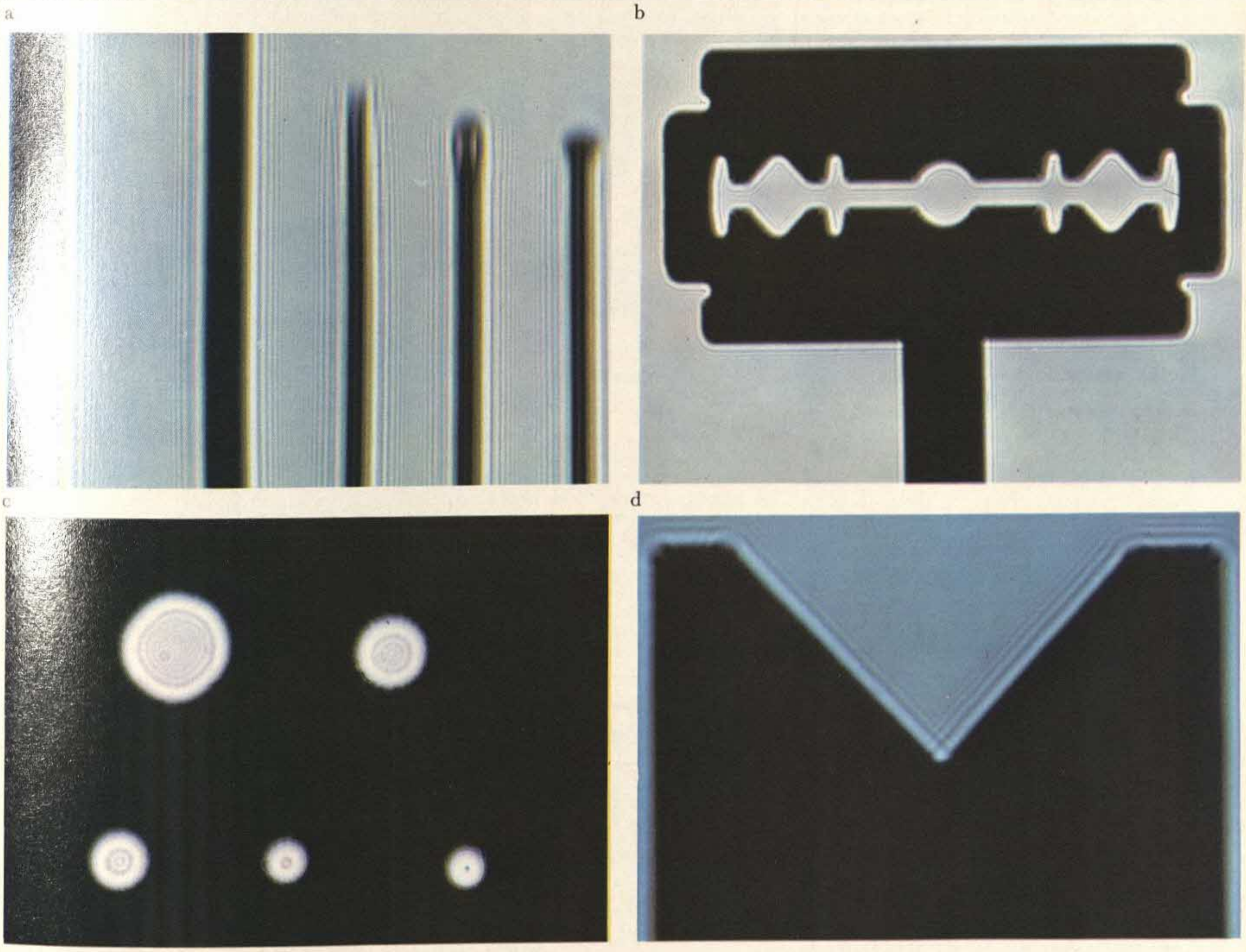
Illustration 4b shows what happens to the light after it strikes the slit. After passing through the slit, it fans out—but not uniformly. It is stronger in certain directions and weaker in others, or even nonexistent. To interpret this phenomenon, it is again necessary to apply the generalized form of Huygens' principle.

DIFFRACTION ON THE EDGE OF A BLADE—Without going into great detail, it is possible to explain the diffraction diagram that resulted when the blade blocked the light waves (as shown in Illustration 2). Illustration 6 shows the wave front that strikes the cutting edge of the blade at intervals. This wave front sends out light disturbances in all directions. However, from the moment at which the wave front is cut by the blade, not all the disturbances can be added together. The sum of the disturbances that pass beyond the blade and reach the screen causes a disturbance that corresponds to the diagram. The calculation of what happens at each point on the surface of the screen, as represented by point **P**, can be made by subdividing the part in front of the blade into segments. Each segment can be used to calculate the strength of the ray directed onto **P** according to the indications of Huygens' principle. Note that at the points at which the light waves are added together and light is produced, the strengths of the light curves are added together. This result is the same as if the two sinusoidal curves were replaced by one of double height. Light perceived by the eye is proportional in strength to the square of the size of the wave. Therefore, a much stronger light can be observed at the points of maximum illumination than that which arrives undisturbed by diffraction.

6



7

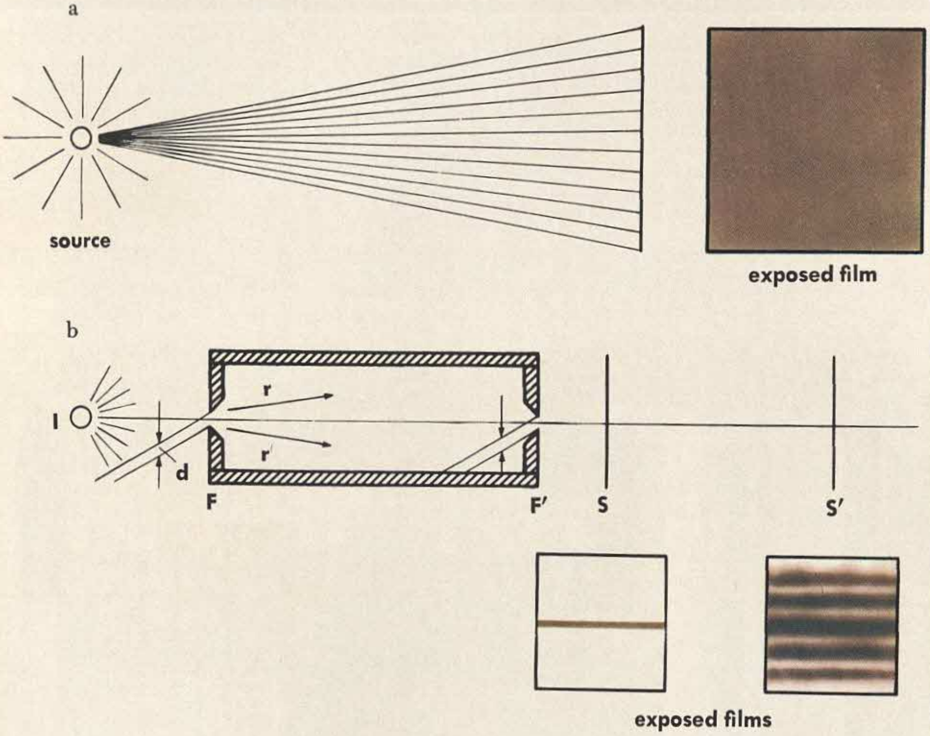


OTHER PHENOMENA OF DIFFRACTION—Illustration 7 shows some of the phenomena of diffraction that can be obtained by using a point of light and collecting on a screen the shadows of various bodies with clear outlines, such as the shadow of a needle (Illustration 7a), a razor blade (7b), holes of different diameters (7c), and a crack with open sides (7d). Diffraction also serves to explain such phenomena as the diffusion of light, to predict the clarity of images that can be obtained by the most perfect optical instruments, and, together with interference, to analyze light by means of a grating.

RESOLVING POWER

the ability to reproduce optical images

1



THE COLLIMATOR—This device consists of a pair of apertures or a lens, which, when aligned along the axis of a light beam, broadens a point source of light to a beam parallel to that axis. The physical form of the device may be either rectangular or circular in cross section, but it must have the aligned pair of apertures or a lens system performing the same function.

Collimators are extensively used to produce beams from radiations that cannot be readily reflected or refracted, either because they emanate from point sources or because they are corpuscular in nature. An ordinary source of light placed close to a sensing or detecting device—such as a photographic plate—illuminates the entire surface with uniform intensity (Illustration 1a). However, if the light source and surface to be illuminated are interposed by a collimator, only a limited area is illuminated (Illustration 1b). The radiation from the source **I** passes through the first aperture **F,** which has a width **d;** the wave front is diffracted to produce a scattering of radiation in directions such as **r** and **r′.** The extent of scattering is determined by the width of the aperture, by the radiation that passes through it, and by the distance of the source from the aperture. Wave fronts that reach the second aperture **F′** are practically flat, so any wave fronts that pass through it are in a beam that is quite parallel.

In the span between the apertures, some diffraction does take place, but the effect is small if the surface to be illuminated is placed close to the second aperture. However, if the surface to be illuminated is moved far from the second aperture to position **S′,** the resolution is greatly diminished. This effect is caused by the propagation of widened images from the second aperture to the illuminating surface. The width of this image is a function of the second aperture width as well as the wavelength λ of the radiation.

The distance **s** between the center of the diffraction image of the aperture and the position of the first minimum of the illumination diagram (see Illustration 5) is given by the equation

$$\sin \alpha = \lambda/\mathbf{s},$$

where α is the angle between the longitudinal axis of the collimator and the line between the center of the diaphragm and the center of the first diffraction minimum. The angle is usually expressed in radians.

The greater the distance between the illuminated surface and the first diffraction minimum, the greater is the linear enlargement of image produced by the beam from the collimator. If an aperture has a width of 10μm, and a green light with a wavelength of 5,000 Å is used, the resolving power of the collimator is very small:

$$\begin{aligned} \sin \alpha &= \lambda/\mathbf{s} \\ &= 0.5/10 \text{ (microns/microns)} \\ \sin \alpha = \alpha &= 0.05 \text{ radians or 3 degrees} \\ &\quad \text{(because } \alpha \text{ is small).} \end{aligned}$$

However, if x-ray radiation with a short wavelength of about 1 Å is used, the resolving power of the same collimator is greater:

$$\sin \alpha = \alpha = 0.0001/10 = 0.00001 \text{ radians or 22 seconds}$$

2

THE RESOLVING POWER OF THE MICROSCOPE—The resolving power of the optical microscope is much less than that of the electron microscope, because the wavelength of visible radiation is much greater than the wavelength of the de Broglie waves (named for the French physicist Louis Victor de Broglie) associated with electron motion. The optical microscope uses light with a wavelength of about 5,500 Å, and the electron microscope uses de Broglie waves about a thousand times smaller. The resolving power of a microscope is determined by the following function:

$$r = 0.61\lambda/NA$$

where *NA* is the numerical aperture of the microscope objective, and λ is the wavelength of the radiation.

Illustration 2a shows an optical microscope. The greatest practical numerical aperture of the objective is about 1.6, and the wavelength of visible light is about 5,500 Å. Certain specialized microscopes illuminated with short wavelength light, for which the wavelength is only 4,000 Å, provide higher resolving power, but the numerical aperture is necessarily smaller than 1.6. The resolving power of optical microscopes is therefore of the order of 0.2μ, approximately half the wavelength of the radiation that is used.

Illustration 2a′ shows a flavobacterium as it appears under an optical microscope with a magnification of about 2,000 X.

Illustration 2b shows an electron microscope. The optical microscopes use a longer wavelength of light than the electron microscope and their objectives have a higher numerical aperture. Therefore the electron microscope has a much higher power of resolution than an optical microscope.

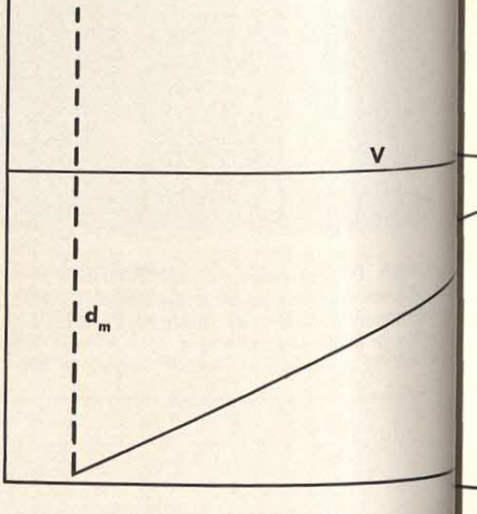
The wavelength of de Broglie waves, which are associated with electron emission, is determined by quantum mechanics calculations. The wavelength is determined by the following equation:

$$\lambda = h/2meV$$

where *m* equals the mass of the electron, *e* equals the value of the charge, *V* equals the voltage with which it is accelerated, and *h* equals Planck's constant. This equation may be simplified to the following form using the value of known constants:

$$\lambda = 12.24/V.$$

3

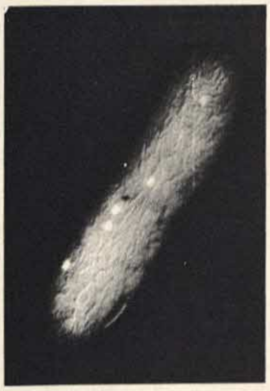


a

a′

b

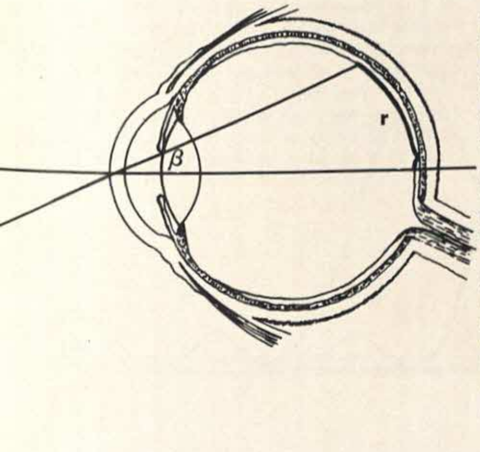
b′



The ability of an optical device to reproduce the details of an optical image is called the resolving power of the device. This ability is limited by the wave-front nature of light and the diffraction effects associated with all such phenomena. All wave-front emanations of energy emerging from a point source are reproduced not as a point, but as a diffraction pattern. Therefore, point sources are resolved or distinguished in their images only if the fringes of the diffraction patterns are sufficiently separated to permit their detection as separate sources. Accordingly, the resolving power of any optical device is indicated by the smallest angular or linear separation of point sources that the device can reproduce.

This measure of resolving power may be applied to all devices that resolve patterns of electromagnetic energy emissions. The measuring technique applies to optical and electron microscopes, collimators, telescopes, and to the human eye—the most sophisticated of all optical devices.

Using this simplified form of the equation, the wavelength of de Broglie waves associated with electron emissions at 1 volt is 12.2 Å; at 10,000 volts is 0.12 Å; and at 100,000 volts is 0.04 Å. Voltages of about 100,000 are used in practical electron microscopes, and wavelengths much shorter than those attained with x-rays are thus obtained. The limitation of the electron microscope is that its objective has a small numerical aperture, usually on the order of 0.01. In spite of this limitation the resolving power of a typical 100,000-volt electron microscope is about 4 Å. High-quality microscopes can provide this resolving power, because they have highly perfected lenses that do not introduce serious aberrations in optical image production. An image produced by an electron microscope is shown in Illustration 2b′—a flavobacterium enlarged 24,000 X.



THE HUMAN EYE—The human eye is the most sophisticated of all optical devices. In calculating the resolving power of the eye, consideration must be given two of its unique features: (1) The diameter of the pupil of the human eye varies with the intensity of incident light, so its resolving power varies. (2) The diameter of the pupil may be as large as 6-7 mm (about 0.23-0.27 in.), so aberrations due to the form of the ocular lens hide the quality of the image and the full aperture cannot be effectively utilized.

The maximum resolving power of the human eye is obtained when the pupil diameter is about 2 mm (about 0.08 in.). When this numerical aperture is used in the equation to determine the resolving power of a telescope, the resolving power of the eye is determined as follows:

$$r' = 12.2/D = 12.2/0.2 = 60 \text{ sec.} = 1 \text{ min. of arc.}$$

The diffraction due to the finite aperture of the pupil therefore causes the diffraction of an image that has an aperture of 1 min of arc (β). The plane of the retina is about 25 mm (about 0.93 in.) behind the ocular lens, and an image with an aperture of about 1′ of arc on this plane has a radius of about 1/200 mm (.000195 in.). Because this is precisely the size of the sensitive elements of the retina, these elements are ideally adapted to the resolving power of the system. The same type of calculations can be used to determine the distance of minimum distinct vision, d_m, which is the minimum distance at which the eye can distinctly perceive objects. This calculation yields a value of d_m = 25 cm (about 9.75 in.). A further calculation indicates that the minimum size object that the eye can perceive is about 0.1 mm (about 0.004 in.).

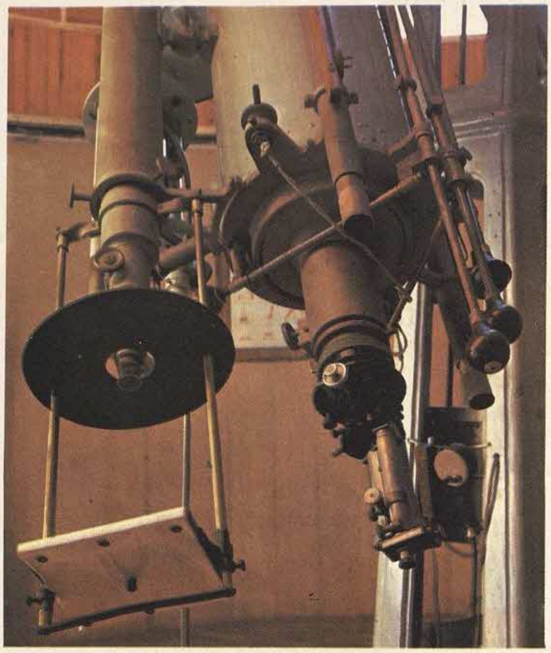
4

THE CASE OF THE TELESCOPE—The angular resolving power of the telescope is given by the simple equation:

$$r' = 1.22/D,$$

where D equals the diameter of the objective in cm and r' is the angular resolving power.

The diameter of the objective of a large reflecting telescope is almost 100 cm (about 39 in.), and the resolving power is approximately 0.1″ of arc. However, this resolving power can only be effectively attained if the wave fronts are transmitted through an atmospheric medium of ideally homogeneous index of refraction. Consequently, the ideal resolving power of a telescope is never attained. Some reflecting telescopes have objectives larger than 100 cm, but they are even more sensitive to atmospheric disturbances that influence the index of refraction of the medium. The illustration shows a telescope used for solar observations from a tower of a large astrophysical observatory in Italy.

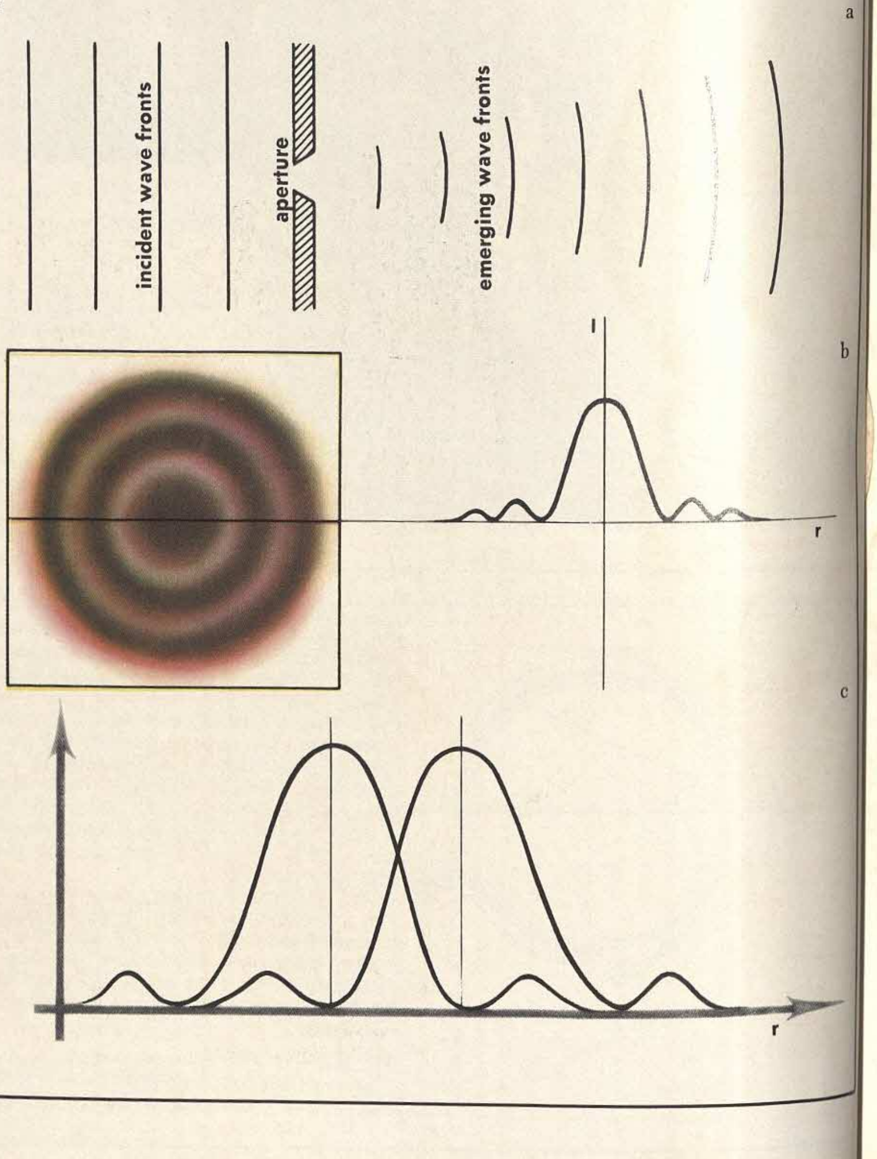


RAYLEIGH'S CRITERION—A criterion used for the assessment of the resolving power of optical devices was developed by the English physicist John William Strutt, Lord Rayleigh, in the late nineteenth and early twentieth centuries. According to this criterion a two-point source of light produces a wave front. This wave front is resolved on an incident wave front if the maximum intensity of one image's center can be distinguished from the maximum intensity center of the other image.

The incident wave front takes the form shown in the lower curve (Illustration 5c), with the intensity shown as the ordinate and the position as the abscissa. If point sources are to be distinguished on a reproduced image, the principal maximum intensity point of one image must not overlap the maximum intensity point of the other image.

In practice, as indicated in Illustration 5a, an emerging wave front, with an intensity distribution as indicated, produces a diffraction pattern. A succession of wave fronts crosses a circular diaphragm, is diffracted, and emerges as a new succession of wave fronts that no longer have plane surfaces. The appearance of the disk with concentric rings is shown in the left-hand part of Illustration 5b, and the curve relating the intensity of incident light to distance from the center of the disk is shown in the right-hand part of the same illustration. The resolving power of a practical optical device is a function of the size of diffraction rings created in a field of vision, or focal plane, by virtue of light diffraction of the objective.

5



LENSES AND ABERRATIONS

imperfections of image

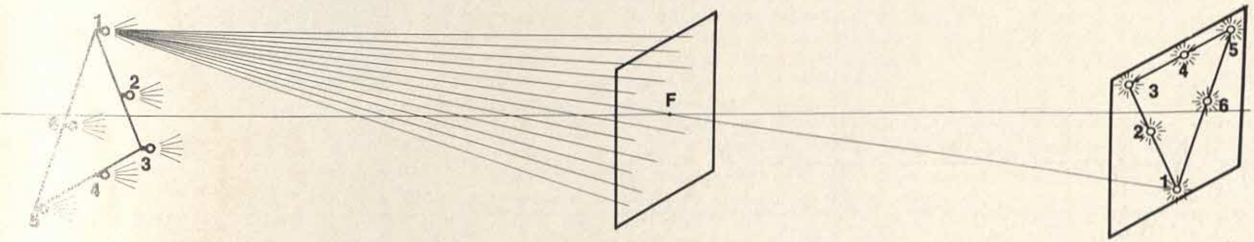
Long before the word Hollywood became synonymous with the illusory world of motion pictures, the "magic lantern" and the static images it projected fascinated viewers. Perhaps it did not annoy those viewers too much that the image on the screen often bore only a scant resemblance to reality, a shortcoming caused largely by what scientists call aberrations.

The light from a source such as an electric lamp is radiated in all directions, illuminating surrounding objects. The light cast on a white screen appears brighter as the screen is moved nearer the lamp. If a lens is placed between the lamp and the screen at an appropriate distance from the screen, an image of the light source is projected onto the screen. This principle is used in photography for the projection of fixed images (photographic slides, for example) as well as moving ones (motion pictures). Due to aberrations, however, what appears on the screen is not a perfect image but a distorted one. Because different lenses will produce different kinds and degrees of aberration, one of the concerns of lens optics is to find the best possible shapes and combinations of lenses to minimize aberrations. This concern presupposes a knowledge of the various kinds of aberration.

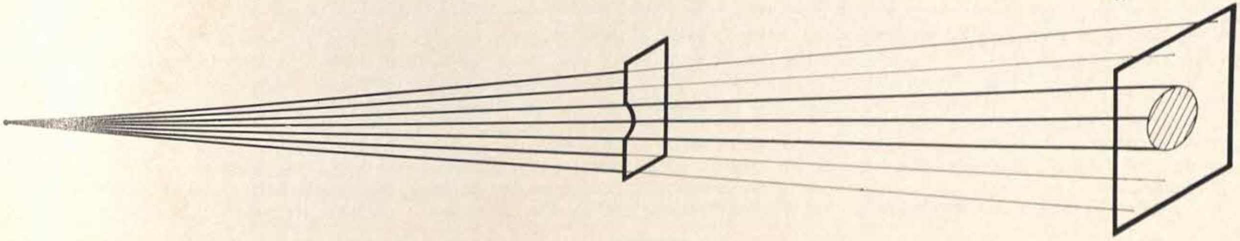
AN IDEAL OPTICAL SYSTEM—Paradoxically, the best optical system happens to be the simplest one—namely, a small hole in an opaque surface. Illustration 1a demonstrates the working principle of the pinhole, the simplest way of producing images. The light source (left) consists of very small lamp bulbs arranged on a support in a triangular pattern. A piece of cardboard with a pinhole aperture **F** in the middle is placed in front of the triangle of light. An opaque white screen is placed on the opposite side. The light from each of the small bulbs radiates in all directions but is blocked by the cardboard, except for those rays that pass through the pinhole and illuminate the screen. Six small spots of light appear on the screen, in the same triangular pattern as the source, but upside down.

A detail of the pinhole optical system, with the size of the hole exaggerated, is shown in Illustration 1b. The larger the pinhole and the greater the distance between it and the projecting screen, the more blurred and dim is the image on the screen.

1a

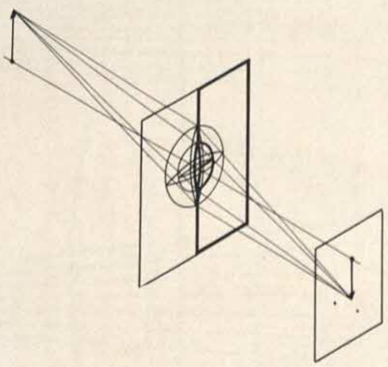


1b

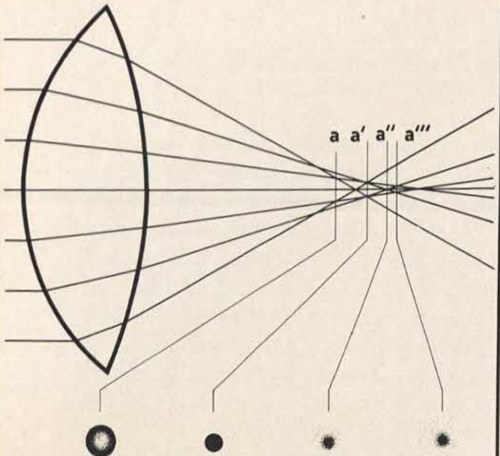


2

ADVANTAGES OF A THIN LENS—In this illustration the pinhole has been enlarged and a biconvex, converging lens with a focal length appropriate to focus the image on the screen has been mounted in the opening. All the rays from a single point source (such as one of the small light bulbs) striking the lens are now bent by the lens and made to converge into a single point on the screen at the right. In this manner many more light rays are concentrated on the screen than in the case of a simple pinhole, and the image is no longer merely a pale shadow of the source, but a very bright representation. However when the lens becomes disproportionately large in relation to its focal length, the images have a number of aberrations and are thus imperfect.

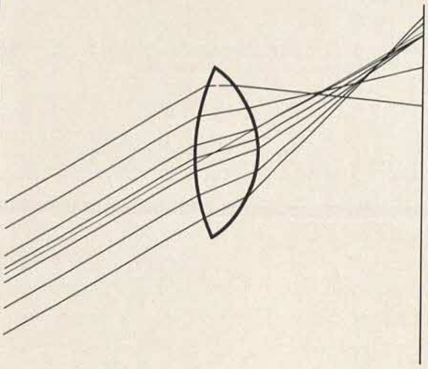


3

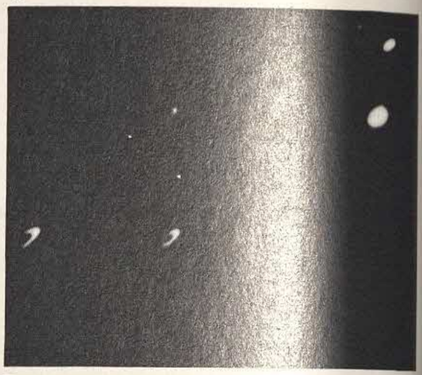


SPHERICAL ABERRATION—Spherical aberration is produced when the rays from a point source of light are not made to converge into a single point, but different parts of the lens converge their incident rays into different points. The lens shown in the diagram is more convergent near its edges than at its center. Blocking the beam of the more convergent rays at **a, a', a''**, and **a'''**, produces out-of-focus images of the point source of light at considerable distances from the lens. Although the image at **a'** has the smallest diameter, it is by no means perfectly focused. Nevertheless, it is referred to as a minimum-blur image. Spherical aberration can be minimized by designing the curvature of one face of the lens so as to compensate for the aberration caused by the other face. Another very practical and effective method of correction is described in Illustrations 5a and 5b.

a



b



c

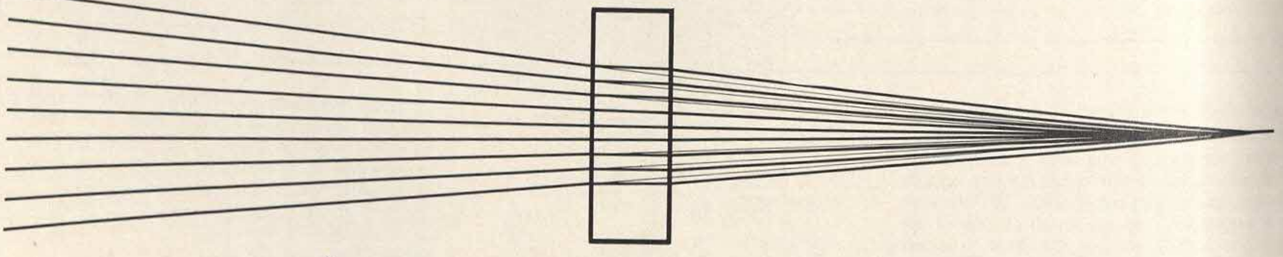
COMA—Spherical aberration is less complex when a lens is struck by a beam of light coming from a point on the axis of the lens than when it is struck by an oblique beam of light, as shown in Illustration 4a. The latter situation is complicated because the focal length of the lens is not unique, but different for various parts of the lens. In this case, the image of a point of light is blurred asymmetrically, appearing as a spot with a tail—roughly the shape of a comet; hence, the name coma for this kind of aberration.

Illustration 4b is a photograph of part of the night sky taken through an optical instrument producing a coma type of aberration at the edges of its effective field. The images of the two stars on the left resemble comets. This does not occur in Illustration 4c, which shows the images of the stars in the middle of the field of the same instrument. Coma aberration is found only away from the axis of the lens, hence, the images of the stars in Illustration 4c are round.

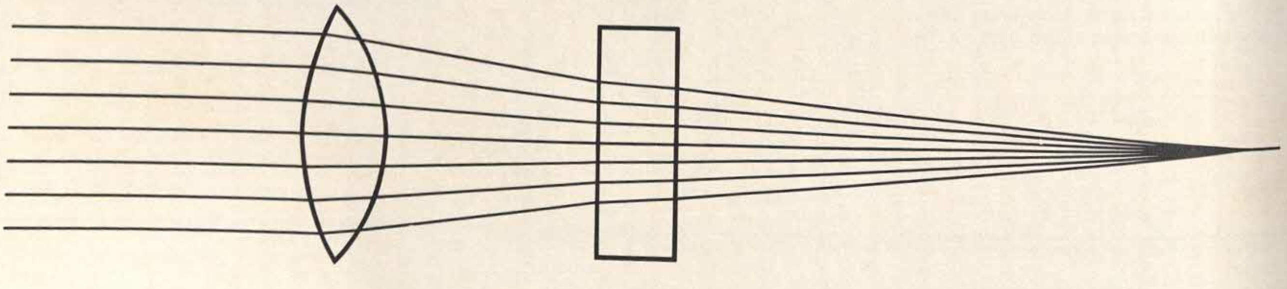
5

COMPENSATING FOR SPHERICAL ABERRATION—Illustration 5a shows how different rays pass through a sheet of glass with parallel plane surfaces. Rays striking the glass at any angle always emerge parallel to their direction of incidence; but with a greater lateral displacement as the angle at which they strike the glass increases.

In Illustration 5b, a thick sheet of glass has been placed behind a converging lens having a high degree of spherical aberration. Without the glass plate, the rays would converge on different points because of the aberration. With the glass present, the peripheral rays, which would otherwise be focused at a point closer to the lens, converge farther away. Spherical aberration is thus partly compensated for by the displacements caused by the glass. Similar arrangements of glass and lenses are used in many optical instruments, such as for example, in ordinary prism binoculars, in which the beam of converging light coming from the objective lens has to pass through the glass of the reflecting prism. This arrangement reinverts the image that was originally inverted by the objective lens, and at the same time compensates for spherical aberration.

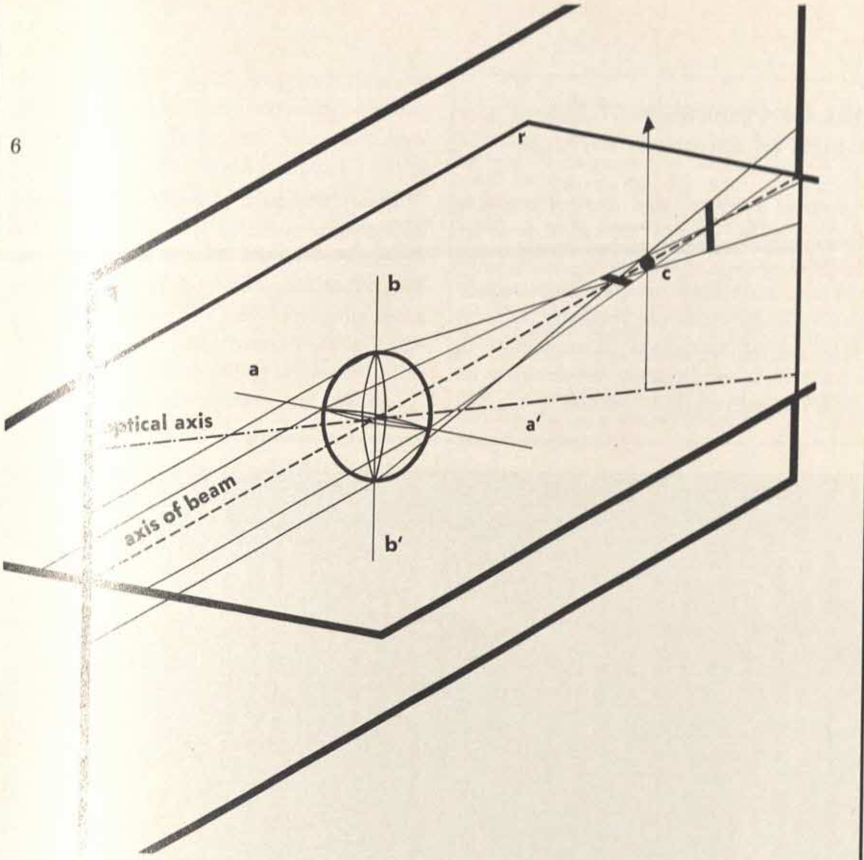


a



b

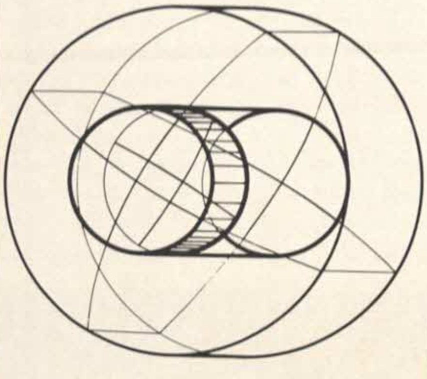
6



ASTIGMATISM—When a beam of parallel rays strikes a biconvex lens obliquely with respect to the axis of the lens, the rays that lie in a plane **q** common to the axis of the beam and the axis of the lens may be focused differently from those in a plane **r** perpendicular to **q** and passing through the axis of the beam. After it has passed through the lens, the beam narrows to form a cone, but the cone does not have a perfect point. The converging rays form first a distorted rod-shaped image perpendicular to q as shown in the diagram. The image assumes this shape because the amount of convergence is greater in the plane **q** than in any other nonparallel plane. After this, the rays that had struck the lens along the horizontal diameter **aa′** continue to converge, while those which struck it along the vertical diameter **bb′** are now diverging from the rod-shaped image, the focus they have reached already. The point of minimum blur is **C.** Beyond this point the rays form a second rod-shaped image perpendicular to **r.**

This defect of a lens is known as astigmatism. Astigmatism and coma are minimized by manufacturing optical systems as symmetrically as possible.

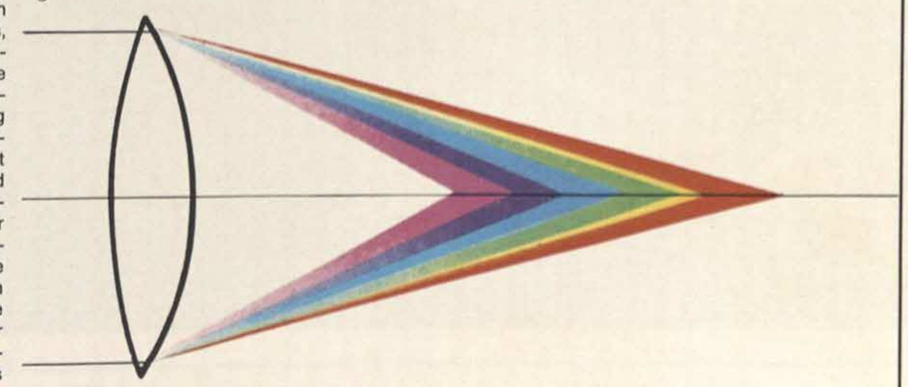
7



APPROXIMATION OF THE THIN LENS—Many of the aberrations described can be practically eliminated if a lens is cut as shown in Illustration 7. Thick lenses that are large in diameter and have short focal lengths are likely to be affected by all the types of aberration described so far. If the diameter of the lens is reduced by cutting out the central part, as indicated, and discarding the rest, the degree of all types of aberration is also considerably reduced. If now the thickness of the lens is reduced too, its focal length will remain unchanged, but all the aberrations, except the spherical, are likely to become insignificant. (Spherical aberration can be diminished by causing the converging beam to pass through a thick plate of glass.) So-called thin lenses are understood to be lenses with practically no aberrations, although such lenses are usually small in size as well as thin. Thin lenses cannot always be used to produce optical systems free of aberrations. If the image needs to be bright, the lenses must have large diameters with respect to their focal lengths. Other aberrations, such as field curvature, are more typical of complete optical systems than of simple lenses.

CHROMATIC ABERRATION—Spherical aberration proper, astigmatism, and coma are known collectively as spherical aberrations. They are associated in the simplest cases with biconvex converging lenses. But spherical aberrations may also occur in converging lenses that are not biconvex as well as in diverging lenses, and even in nonoptical systems such as those concerned with focusing of electrons, protons, and other charged particles by electric and magnetic fields. Optical lenses are also affected by so-called chromatic aberration. The functioning of a lens as a focusing device is based on the deviation in the direction of light rays when they are refracted at the surfaces that separate air from glass and glass from air. This specific amount of deviation depends on the wavelength of the color of the light. If the light is white (being a mixture of all the colors of the spectrum), the various colors will focus as shown. On a screen placed at different distances from the lens, the lens will focus the image in a different color each time, depending on which color component of the white light happens to be in focus at the particular distance. The image of a white dot will then be seen as a colored dot encircled by an out-of-focus, diffuse ring of a complementary color. Which two colors appear in the dot and its halo respectively depends on the distance from the lens to the screen. An image formed by a lens affected by chromatic aberration has a rainbow-colored edge and is blurred. The most practical approach to correcting chromatic aberration is the arrangement of lenses in pairs so that each lens will cancel the chromatic aberration of the other.

8



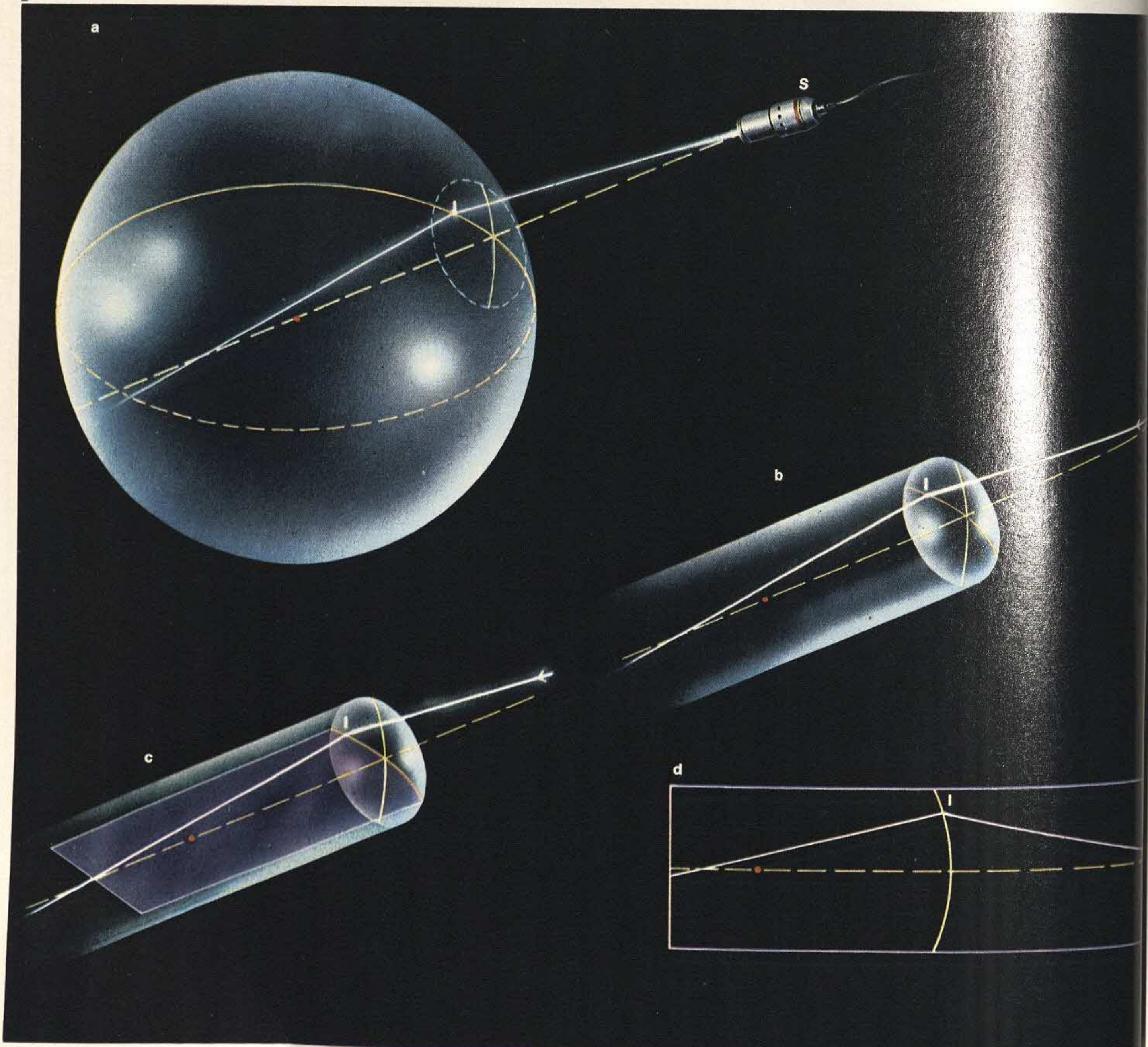
DIOPTRICS | the fundamentals of the optics of spherical surfaces

Scientists of antiquity studied the reflection of light from spherical and other surfaces. They knew the fundamental law of reflection: the angle of incidence equals the angle of reflection (where both angles are measured with respect to the normal). Moreover, they used this principle to address problems of reflection from spherical mirrors, and they attained a considerable degree of sophistication. Legend credits Archimedes with using mirrors to set fire to Roman ships attacking Syracuse. However, the ancients knew little about refraction at spherical—or other—surfaces, because the laws of refraction had not been defined.

Today, the path of a light ray through any system of lenses can in principle be determined by the laws of refraction, but the calculation of an optical path through a complicated series of lenses is complex—and some photographic objectives consist of as many as ten lenses.

Lenses are transparent bodies, gen-

1



erally made of glass (but sometimes of plastic materials) and bounded by two spherical, parabolic, or other surfaces. Spherical lenses require special consideration. The problem presented in the accompanying illustrations is this: what path is followed by a light ray as it travels through a lens (or lenses) and crosses spherical surfaces? Every surface dividing two materials of different indices of refraction causes the refraction of a light ray.

THE NATURE OF THE SPHERICAL REFRACTING SURFACE—Illustration 1a represents a sphere of refracting material. It may be made of air or water, but glass is the substance usually considered in optics. A ray of light coming from source **S** is made to fall on this sphere at point **I.** This ray has been collimated either by means of a small projector or by making it pass through some diaphragms, but these details are not shown. Dioptrics studies the path of such a ray.

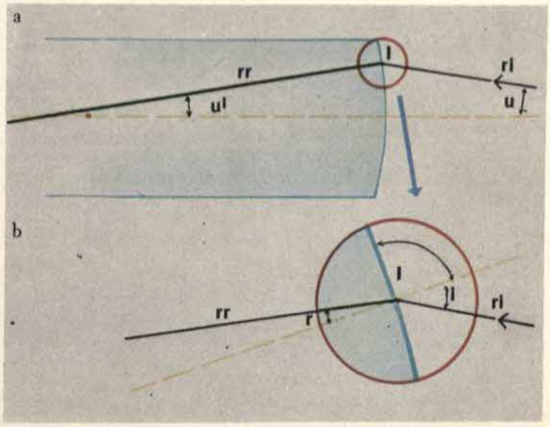
In Illustration 1b the problem is simplified by neglecting parts of the sphere that are not used. A straight line is drawn from the point source of light to the center of the sphere and a cylinder is constructed around this line. The base of the cylinder is spherical and contains point **I.** The cylinder need only be continued to be long enough to contain the entire path of the light ray through the original sphere.

In Illustration 1c both the ray of light that has undergone refraction and the line joining the light source to the center of the sphere are located in the same plane. Hence, if the cylinder is cut along the plane, the section obtained includes the path of the light ray. The circular profile of the sphere and the axis between the source and the center of the sphere are all the geometric elements that are needed to calculate the path of the ray.

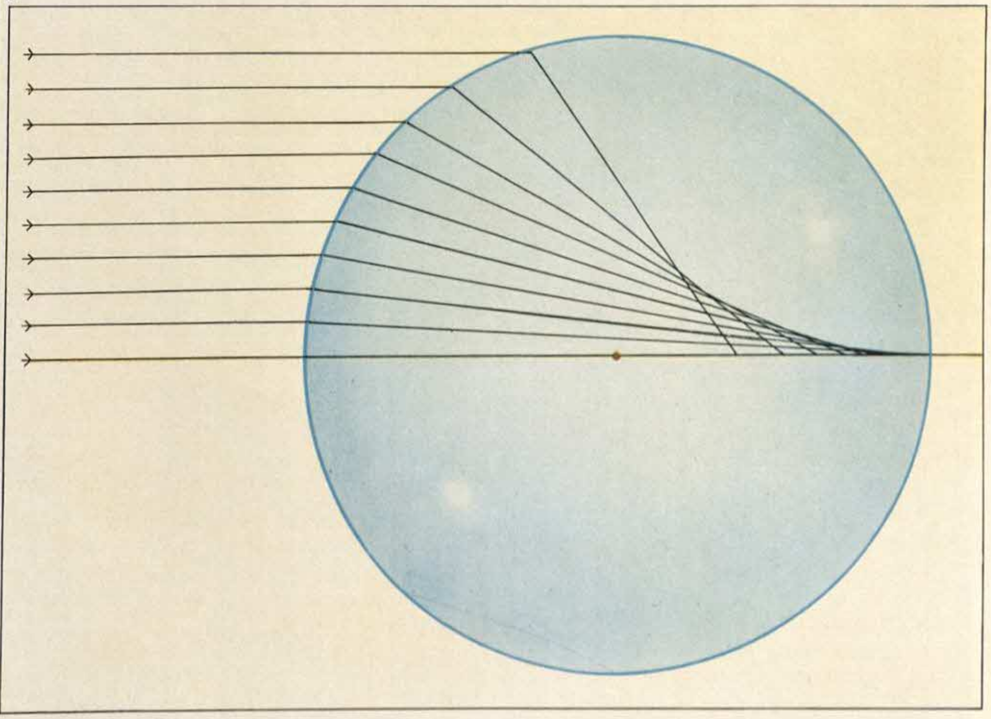
Illustration 1d shows the geometric scheme for the refraction of a light ray. However, the original three-dimensional figure should be kept in mind.

THE PROBLEM OF DIOPTRICS—In Illustration 2a a ray of light is shown falling on a sphere. To calculate the path of this ray inside the sphere and the point at which it intersects the axis, an enlargement of the encircled area is shown in Illustration 2b. Here, the spherical surface appears essentially as a plane. Hence, the law of refraction can be applied: sin i/sin $r = n$, where n is the index of refraction of glass. The refracted ray is seen to follow line **rr.** To calculate where line **rr** intersects the axis, some other parameters must be known—the distance of **I** from the axis, the radius of curvature of the surface, and the angles **u** and **u′** between rays **ri** and **rr** and the axis.

2



3



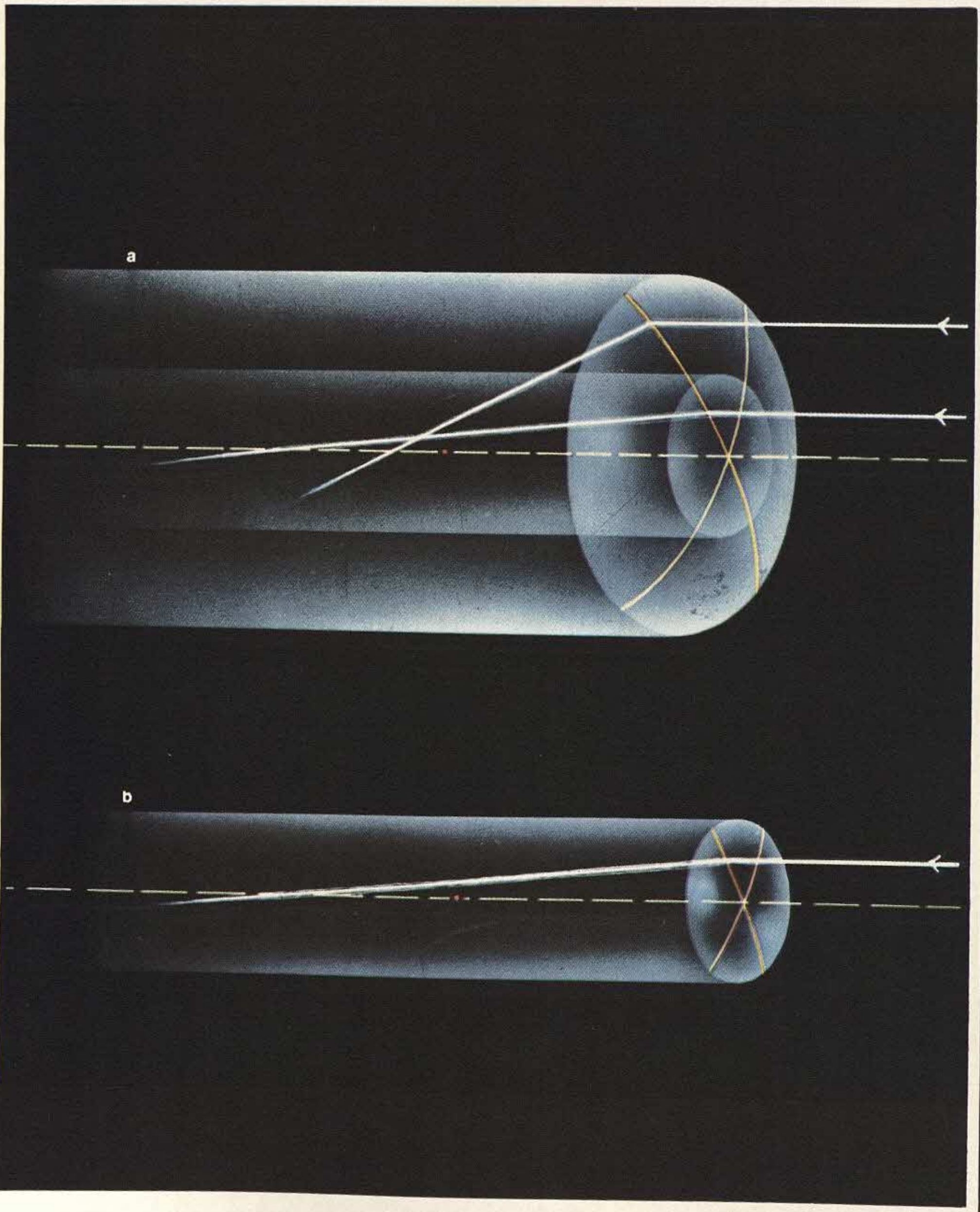
A PARTICULAR CASE: PARALLEL RAYS—If parallel light rays are incident on a sphere, some rays are farther from the axis than others. Rays that are close to the axis (called paraxial rays) all converge at practically the same point, but rays that are some distance from the axis converge at different points. The point at which paraxial rays converge is called the focus of the refracting surface. The name *focus* (from the Latin word for "fire") was given to this point because sun rays, if made to converge in this manner, can set an object on fire at this point.

One definition makes it possible to measure and calculate refraction. Light travels more slowly in water than in air; the speed of light is even less in glass than in water.

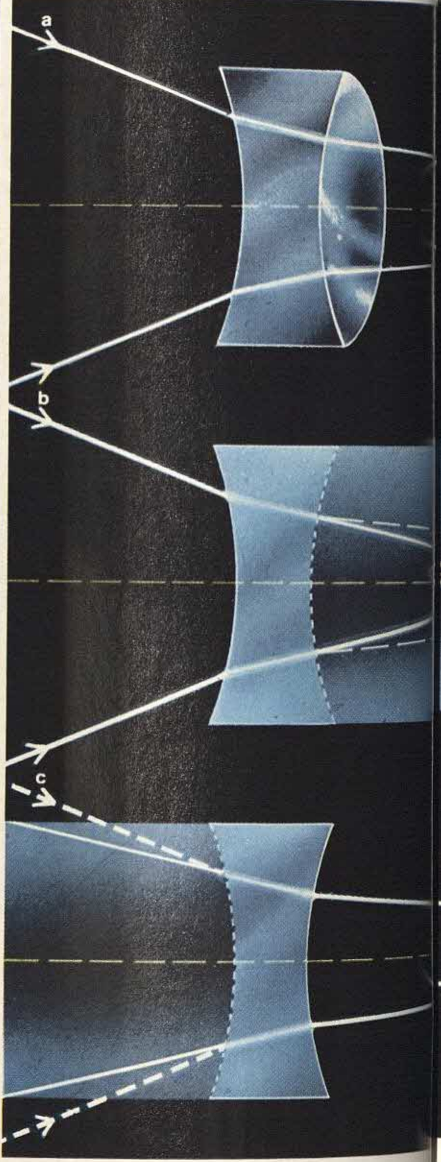
The ratio of the speed of light in air to that in any other medium is the refractive index (or index of refraction) of the substance. The refractive index can be found for any substance by comparing the angle of incidence with the angle of refraction; the ratio between the sines of the two angles is the refractive index. Likewise, the angle of a refracted beam of light can be calculated from the incident angle and the angle of refraction.

PARAXIAL RAYS—An important case is that in which all the rays are parallel and close to the axis; they, therefore, converge at the same point. This case can be considered by reducing the surface from that shown in Illustration 4a to that shown in Illustration 4b. All paraxial rays impinging upon this reduced area converge at the focus. The focus can thus be found by calculating the path of only one ray. In some cases, however, this approximation is not practicable. In the design of a photographic objective, for example, a long series of mathematical calculations must be carried out to solve this problem.

4

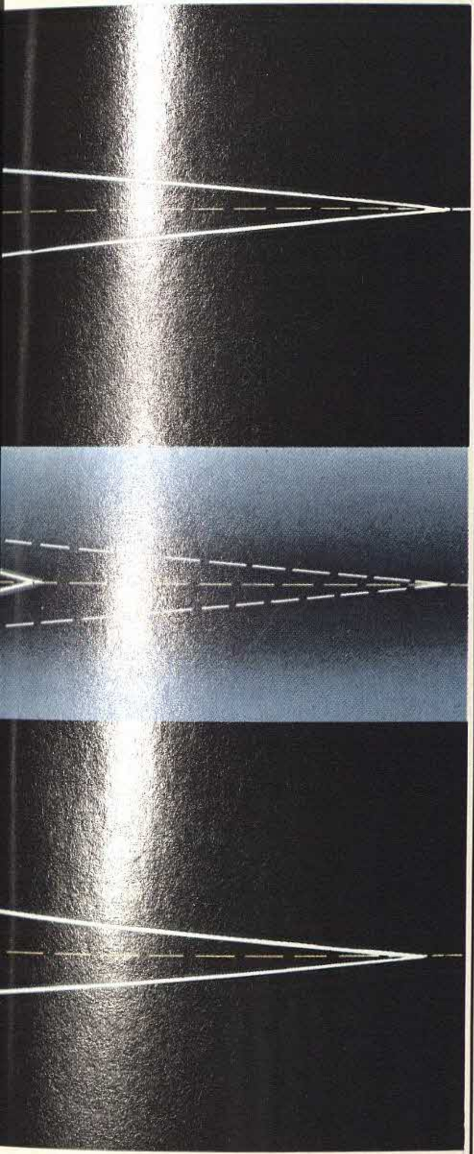


5



A LENS: A COMBINATION OF TWO REFRACTING SURFACES—Illustration 5a shows a diverging lens on which the strongly converging beam of light is impinging. To calculate the paths of the rays after they have passed through the lens, separate consideration of the two surfaces is sufficient—presuming, of course, that the curvatures are small enough that the incident rays can be considered paraxial.

Illustration 5b shows the paths of the rays that have crossed the first refracting surface.



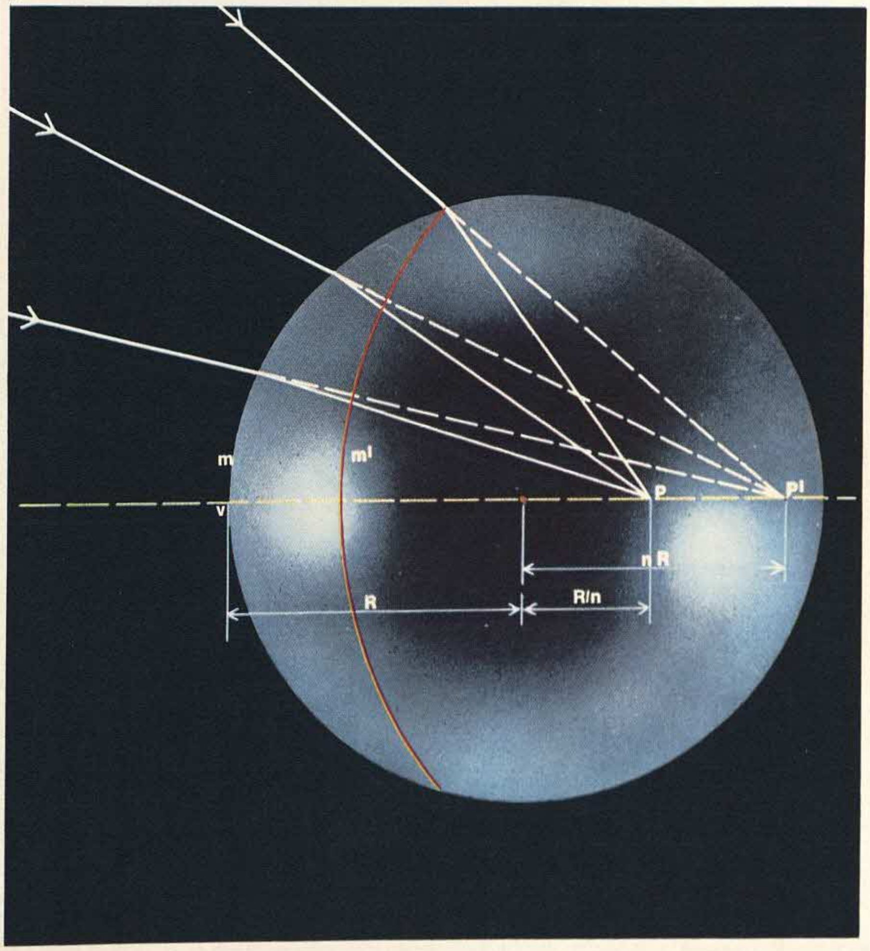
The rays travel in a medium independent of any future refractions or part refractions. This is the reason the refractions can be considered independently. The dotted lines represent the path that the light follows.

In Illustration 5c the rays encounter a second surface. Only the angle of incidence is important, and the past history of the rays is irrelevant. The rays could have always had this inclination (represented by the solid lines on the left). The paths of the rays on passing through the second surface are shown, right.

THE APLANATIC POINTS OF A SPHERE— Parallel rays passing through a spherical surface do not converge at the same point unless they are very close to the axis. This illustration shows a converging of rays that are not parallel. The rays incident from the left would all converge at **P′** if they did not refract on encountering the sphere. However, they converge at **P** due to refraction. If the distance of point **P′** from the vertex **V** of the spherical surface and the radius **R** of the sphere have arbitrary values, the point of convergence **P′** is only approximate; some of the rays converge a little nearer to the vertex of the spherical surface, while others converge a little farther away. However, if the distance from the vertex **V** to point **P′** is equal to **R + nR** (where **n** is the index of refraction), then the distance from **V** to point **P** is given by **R + R/n** and all the rays rigorously converge on this point. **P** and **P′** are called the aplanatic points of the sphere and suggest a very important application to the construction of lenses: the aplanatic meniscus. This is used in the construction of objectives for high-powered microscopes.

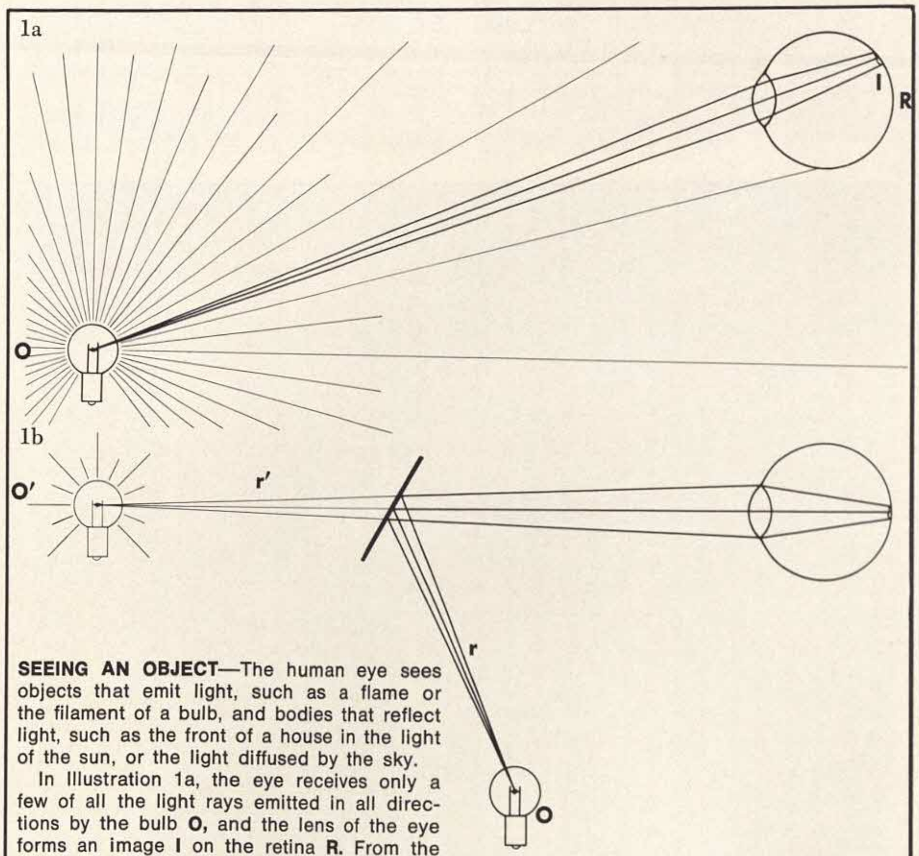
In this illustration the meniscus consists of a lens delimited by two spherical surfaces; the first of these **m** is identical to the surface just considered, while the second has its center of curvature at the point **P** and is shown in the diagram by means of the full red line. The rays do not undergo any deviation when they meet the second surface because all of them are impinging on it normally; the rays continue as if they were traveling inside the refracting substance of the meniscus. The front lens of certain objectives is constructed to utilize these aplanatic points. The rays can thus be made to converge toward the eyepiece.

6



MIRROR OPTICS | the geometry of reflection

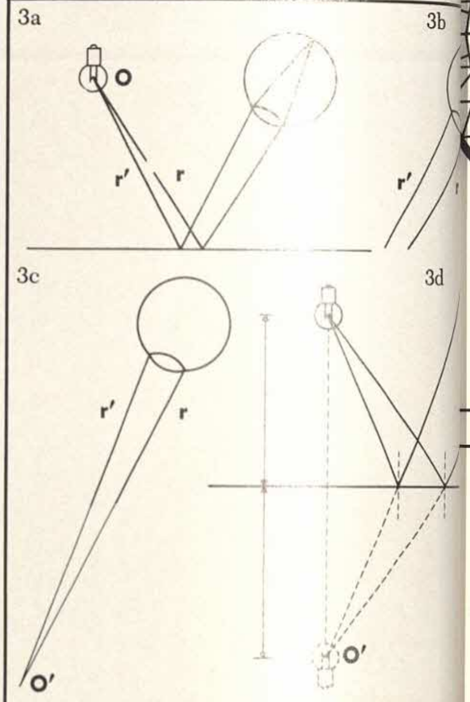
Light is electromagnetic radiation that can be detected by the eye. Much can be learned about light rays, which travel in straight lines until they meet an obstacle, without actually taking their electromagnetic nature into account. The branch of optics in which the wave nature of light is not considered is called geometrical optics because the optical laws governing it are formulated in geometric terms. (When the wave nature of light is taken into account, the subject is called wave or physical optics.)



SEEING AN OBJECT—The human eye sees objects that emit light, such as a flame or the filament of a bulb, and bodies that reflect light, such as the front of a house in the light of the sun, or the light diffused by the sky.

In Illustration 1a, the eye receives only a few of all the light rays emitted in all directions by the bulb **O**, and the lens of the eye forms an image **I** on the retina **R**. From the position of this image the brain deduces the direction from which the object has been seen. The functioning of this surveying mechanism is based on the perception of the direction from which light rays originate.

Illustration 1b shows how the human eye can be deceived by a mirror placed to cause a deviation of the path of light rays. The light rays seem to be coming from the direction **r′** instead of from **r**, and the bulb appears to be in position **O′** instead of the actual position **O**.



THE PLANE MIRROR—Illustration 3a shows how a plane mirror can be used to form optical images. The bulb **O** emits light rays in all directions. Of all these rays, two, **r′** and **r**, have been indicated as they strike first the mirror and then the eye of the observer.

Illustration 3b shows only the part of the rays **r′** and **r** that the eye sees. The other parts—the mirror, the light source, and the light rays between the source and the mirror—are not shown.

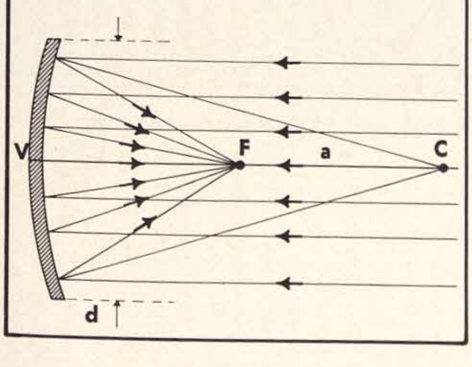
Illustration 3c shows how the eye seems to perceive the rays as emerging from the source **O′**.

Illustration 3d shows in detail how an image is formed by a plane mirror. The light rays appear to originate from a source **O′** behind the mirror. Since the source **O′** does not give off any direct light (although it may seem to do so), it is called a virtual source. The image **O′** is, therefore, a virtual image.

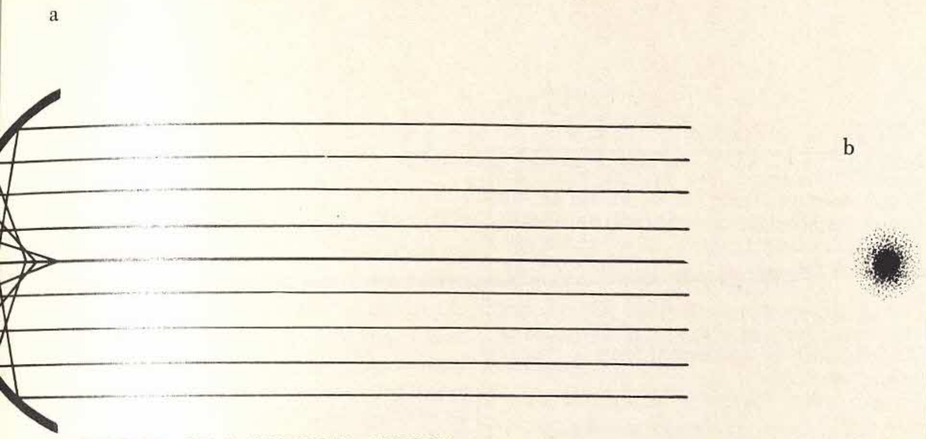
Light rays travel in straight lines only when the medium through which they travel is of the same optical density throughout. Interesting phenomena occur, therefore, at the interface of two different media. For example, when light rays pass from air to water (which has a greater optical density than air), the light rays are both reflected and bent (refracted) at the surface of the water. The light rays then continue to move in straight lines until their paths are modified by other obstacles.

2

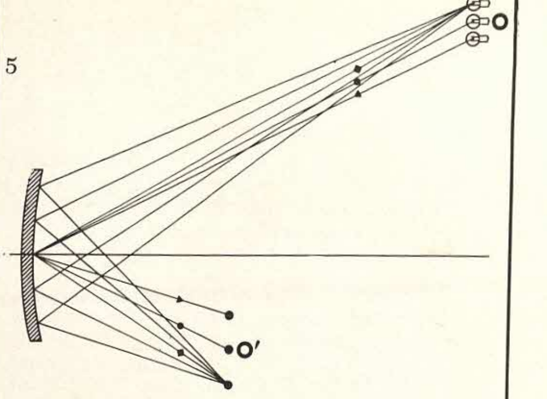
THE SPHERICAL MIRROR—The illustration shows how a beam of parallel rays of light is reflected from a mirror to meet at the focus point **F. V** is the vertex or axis of the mirror. If the mirror has a diameter **d** that is not too large relative to the distance **VF**, light rays directed parallel to the axis onto the mirror will converge at point **F**, which is midway between the vertex **V** and the center of mirror curvature **C**.



The study of light rays falling on a flat surface is much less complex than the study of reflection and refraction of light from a curved mirror. However, understanding the theory of the formation of images by reflection in curved mirrors is important in designing instruments that use mirrors instead of lenses.



5



ABERRATION OF A SPHERICAL MIRROR—A spherical mirror converges the parallel rays that strike it at its focus only when the diameter of the mirror is small in comparison with its radius of curvature: that is, if the mirror has a limited spherical section. In Illustration 4a, however, a wide spherical section is shown in order to illustrate what is commonly known as spherical aberration. The rays that strike the mirror converge at different distances from the focus. Those striking the periphery of the mirror converge at a different point from those striking the center. A figure representing an image affected by spherical aberration is shown in Illustration 4b; the image is not sharp but blurred.

THE FORMATION OF IMAGES—Illustration 5 shows how a spherical object forms images. Three sources of light **O** are placed in front of a spherical mirror. An image of each of the sources represented by **O** is formed in the position **O′**. This occurs only if the diameter of the mirror is small in proportion to its radius of curvature.

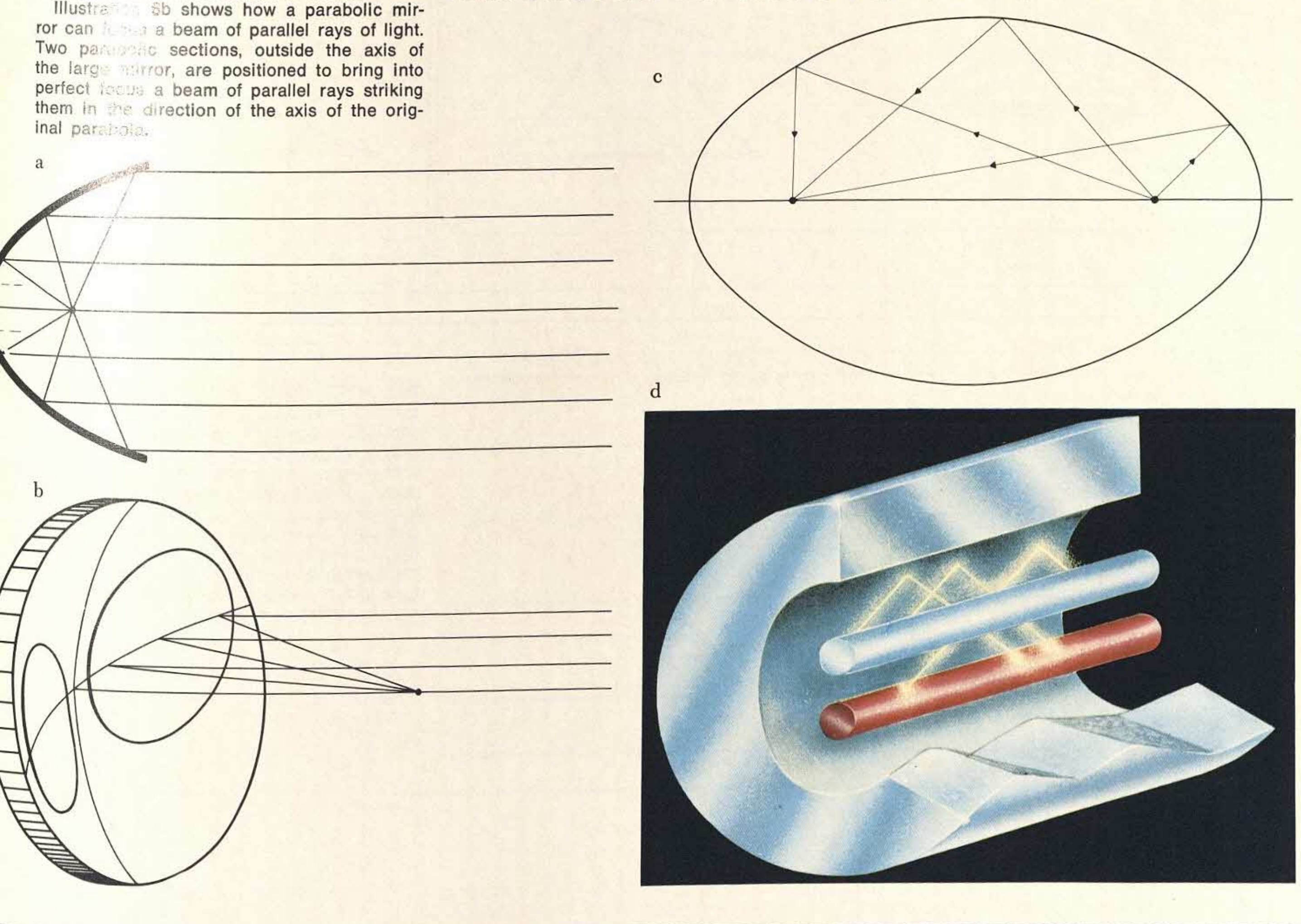
6

A PERFECTLY STIGMATIC MIRROR—A parabolic mirror (Illustration 6a) is stigmatic: it focuses all parallel rays at one point and reflects those originating at the focus as parallel rays. The more an optical surface deviates from the spherical, however, the more difficult its construction becomes. Parabolic mirrors are, therefore, usually made with narrow apertures, indicated by the arrows.

Illustration 6b shows how a parabolic mirror can focus a beam of parallel rays of light. Two parabolic sections, outside the axis of the large mirror, are positioned to bring into perfect focus a beam of parallel rays striking them in the direction of the axis of the original parabola.

Illustration 6c shows how an ellipse is analogous to a parabola. An ellipse, rotated around its major axis, forms a surface of revolution that brings into focus, at one of the focal points, the rays that emerge from the other focal point. In a few rare cases ellipsoidal mirrors are constructed to bring radiation between finite points into focus. This method is rare, because of the difficulty of obtaining ellipsoidal surfaces.

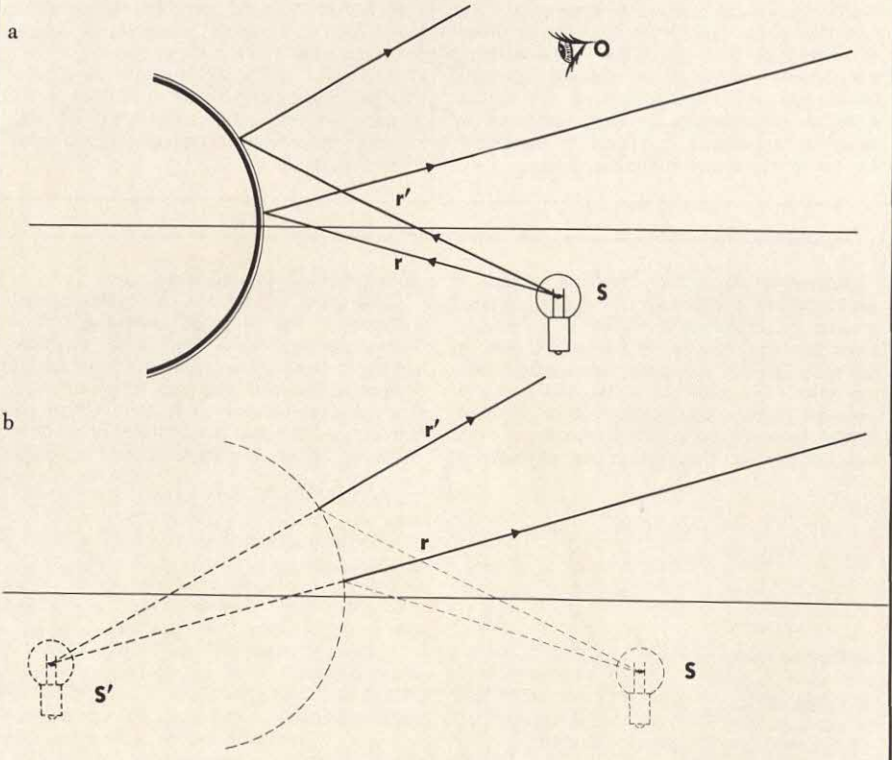
Illustration 6d shows how an important property of the ellipsoid is exploited in a solid laser device. A flash tube containing xenon or some other suitable gas is used to illuminate the ruby rod that lights the laser. A cylindrical reflector with an elliptical section is used to focus the light on the synthetic ruby rod, which is parallel to the flash tube.



7

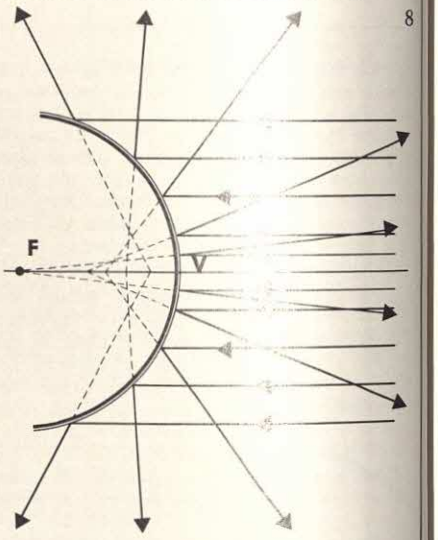
THE CONVEX MIRROR—Illustration 7a shows how images are formed on a convex mirror. A light source **S** emits light in all directions. Part of this light strikes the mirror; **r** and **r′** represent two rays that strike it and are reflected; **O** represents the eye of the observer.

The source of light and the mirror have been omitted from Illustration 7b, leaving only the path of the rays **r** and **r′** as reflected by the mirror. The rays, when they are reflected in the mirror, are diverged from their initial direction. Their origin (as far as the eye of the observer is concerned) appears to be the nonexistent source **S′**. When the eye perceives the rays **r** and **r′**, it cannot know whether they come directly from source **S** or from **S′** by reflection. In fact, the observer has the impression that a source **S′** (indicated in the illustration by faint dotted lines) is directly behind the mirror.

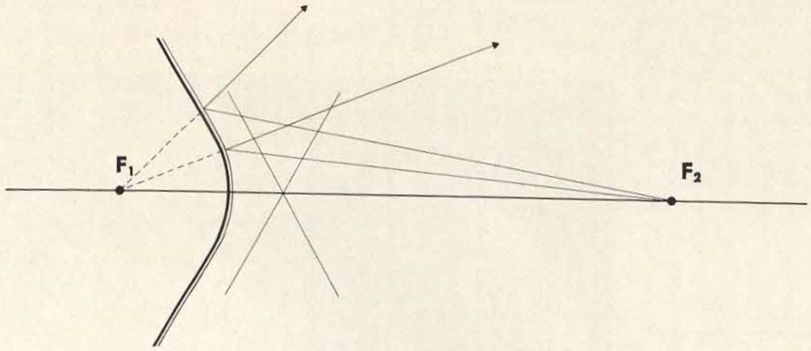


SPHERICAL ABERRATION OF CONVEX MIRRORS—Convex mirrors also have their aberrations. Illustration 8 shows a beam of parallel rays striking a spherical mirror. Those that strike near the principal axis converge at the focus **F,** while those that strike at points distant from the axis converge at a point nearer the vertex **V.** This spherical aberration (nonfocusing of parallel rays at the principal focus so that a distorted image is produced) is analogous to that of a concave mirror.

8



9



THE HYPERBOLIC MIRROR—One type of concave mirror, the parabolic, is perfectly stigmatic (produces geometric points as images of geometrical points) to a beam of light coming from infinity. Convex mirrors are never stigmatic. Only a hyperbolic mirror has the property of causing a beam of light from one point to converge on another point, but the two points must be the foci **F_1** and **F_2** on the hyperbolic section of the mirror. A concave mirror of this type is used to collect light rays in reflecting telescopes.

OPTICAL ILLUSIONS | experiments in the psychology of sensations

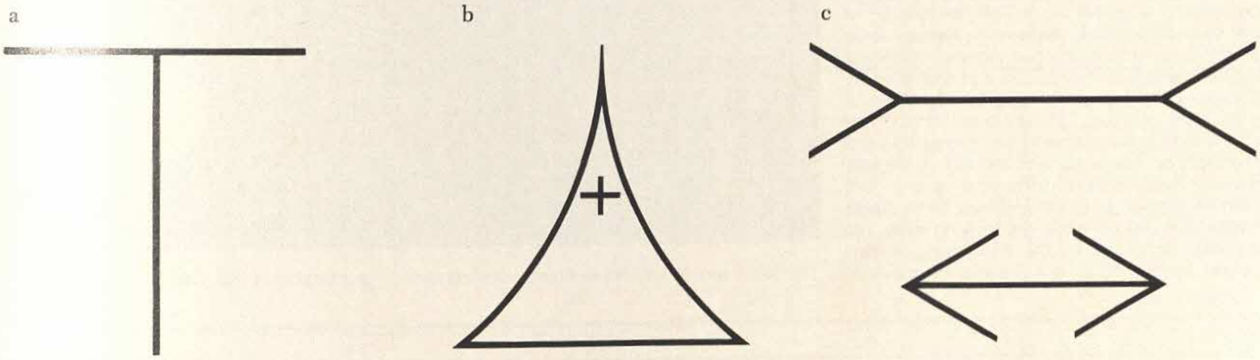
The study of optical illusions opens the door to many profound observations about man and nature. Some interesting and relatively simple experiments can be performed that relate to physics, physiology, and psychology. In these experiments, the tricks that the eye plays on the brain can reveal otherwise hidden aspects of man's external world.

The reasons that observers are often deceived by particular images are many. They are found in such phenomena as light, color, size, dimension, angle, and distance. Everyday experience provides the means, and often the necessary apparatus, to create and explain such phenomena.

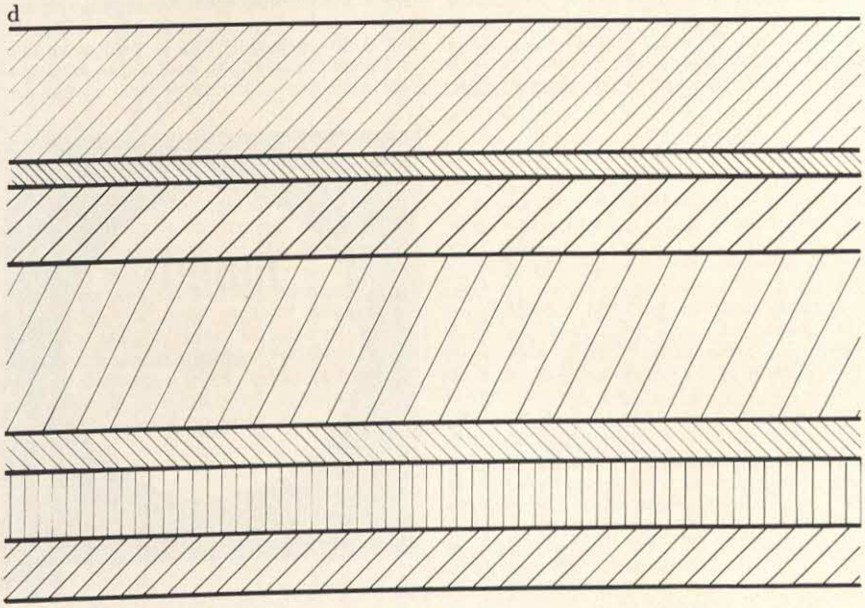
Psychological effects of color are directly related to optical principles. In the dark, for example, a red light appears to be closer than it really is, and a blue light seems farther away. The eye adjusts its focus in a different way for the long-wave red rays and the short-wave blue rays. One of the most interesting reactions of the eye to colors is the afterimage, which can be illustrated by a simple experiment. By staring at a spot of color about an inch square on a piece of white paper and, after about 30 seconds, staring at a blank sheet of white paper, one is able to "see" a dot of color roughly complementary to the original spot.

1

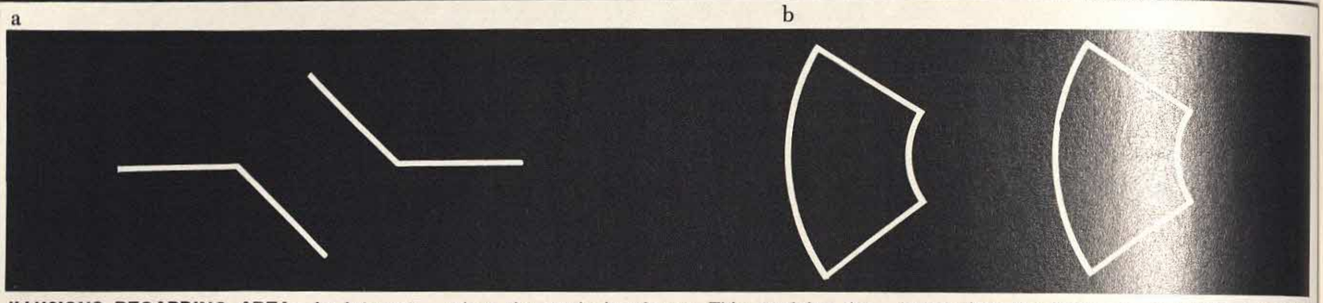
FOUR COMMON OPTICAL ILLUSIONS—Some optical illusions are physiological in nature, and others are purely psychological. In the letter T, for example, the vertical line appears to be longer than the horizontal line because the latter is seemingly divided in two (Illustration 1a). Subconsciously, the brain compares the vertical line with each half of the horizontal line. The cross in Illustration 1b, which is about halfway between the top and bottom, appears closer to the top because of the converging lines.



A line with arrowheads (Illustration 1c) will seem longer if the arrowheads point inward, and shorter if they point outward. The reason for this is that the brain unconsciously includes part of the arrowhead in the length of the horizontal line. However, it does not include the arrowhead whenever it points inward. It is extremely difficult to overcome this sensation and make a correct estimate. The optical illusions shown in Illustration 1d are often found in mechanical drawings. When the space between two parallel lines is hatched with short oblique lines, the eye perceives the principal lines as converging rather than parallel. In fact, the subconscious tends to see the short lines as horizontal. If the dominant oblique lines tend to be horizontal, the parallel lines will seem to incline. Consequently, true parallel lines seem nonparallel despite the fact that they are equidistant at any point.



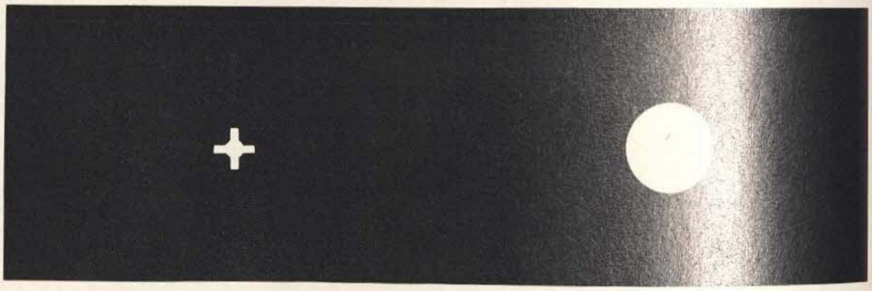
2



ILLUSIONS REGARDING AREA—An interesting optical illusion (Illustration 2a) is that in which the sides of an angle tend to appear inward when the angle is acute, and outward when the angle is obtuse. This explains the seemingly obvious fact that the right horizontal side is lower than the one on the left. In Illustration 2b, the figure on the right appears larger than the figure on the left because the widest area adjoins the shortest area of the figure on the left.

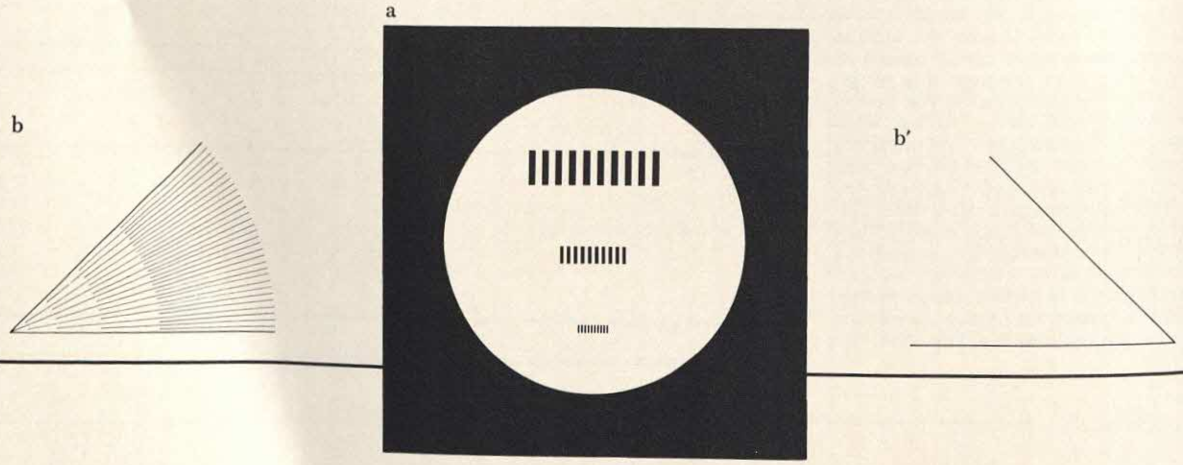
EXAMPLE OF ADAPTATION—This illustration shows how the eye adapts to unfavorable or exceptional circumstances. The "blind spot" in the eye, that point where the optical nerve is attached to the eyeball, cannot perceive light and images. This can be demonstrated by observing the cross with only the right eye at a distance of about 25 cm (about 10 in.). The disk on the right seems rather large. With the eye fixed on the cross, and moving to the left, the disk can be made to appear smaller and even disappear. In this experiment each image (cross and disk) is observed by both eyes, but the part not seen by the blind spot is perceived by the other eye. Even with one eye, this blind spot is seldom realized, which clearly demonstrates the processes of adaptation.

3



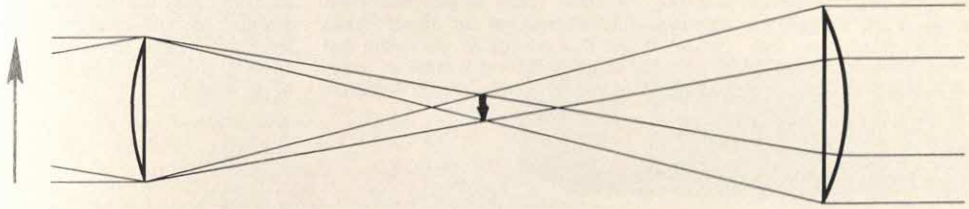
4

IMPORTANCE OF DETAIL — The details in a figure tend to augment its apparent dimensions because the eye perceives many of the details and, in distinguishing them from one another, will make the whole appear much larger. The different lines in Illustration 4a are hard to estimate even though they are 2, 1, and .5 cm in length. This is due to the eye and brain attempting to sort out as many details as it can at the same time. In the two angles shown (Illustration 4b and 4b′), the one on the left seems larger because its entire area contains fine lines of different sizes (details).

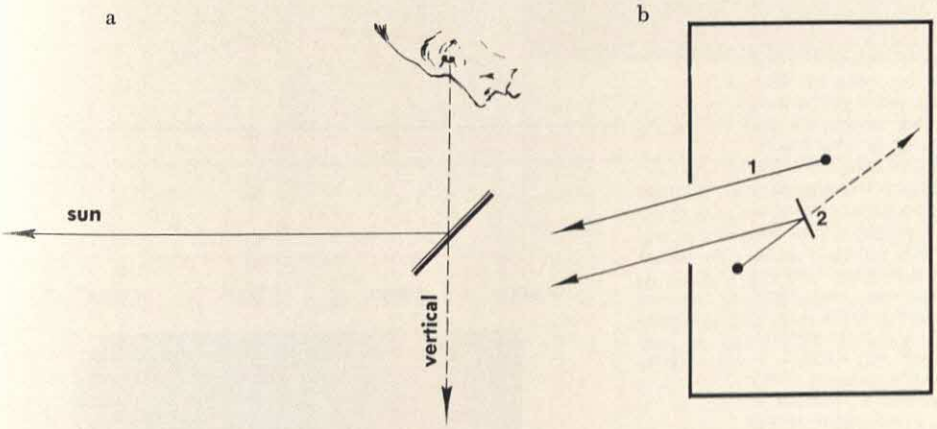


OPTICAL ILLUSIONS IN EVERYDAY EXPERIENCE—The previous illustrations are often used to demonstrate optical illusions, but many more exist in everyday experience. For example, in designing a series of buildings, an architect will never place a tall slender building next to a low squat one, because such a contrast would offend aesthetic sense. In observing objects, the eye tends to exaggerate dimensions in relation to one another more than in the simple designs previously illustrated. Consider, for example, an image seen through the telescope. With the eye close to the eyepiece, all objects seem smaller than they really are. Unconsciously, anything seen through a telescope is considered not to be very big. A simple telescope consists of two lenses and a tube that is twice as long as the focal length of each lens. Those objects seen through this device will seem two or three times smaller than they really are; this illusion becomes even more pronounced since these images are perceived as upside down. By looking at the same object with the naked eye and with the telescope, images of the object are superimposed and they appear identical. When designing instruments that need not enlarge an image (periscopes, for example), manufacturers use a magnification of 1.5 X to counteract this natural tendency to see things smaller than they really are.

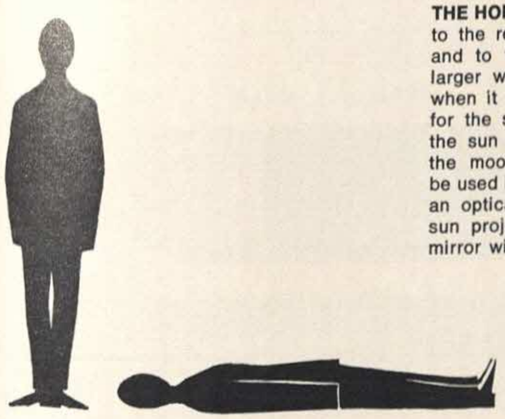
5



6



c



THE HORIZON—Optical illusions are often due to the relationship of objects to one another and to the horizon. Thus, the moon seems larger when it appears on the horizon than when it is high in the sky. The same is true for the sun and certain constellations. Since the sun has the same apparent diameter as the moon and is easier to observe, it can be used in a simple experiment to demonstrate an optical illusion. At sunrise or sunset, the sun projected onto the floor by means of a mirror will appear quite small (Illustration 6a). The image is so small in fact that it is difficult to believe that some of the optical systems (eye, mirror) through which it must pass do not reduce its size.

Similarly, a ship on the horizon provides the basis for another experiment. The angle the ship subtends may amount to a degree, a tenth, or smaller fraction. The same ship viewed in a mirror (Illustration 6b) will appear much smaller and may even be hard to find at all.

Another experiment can be performed easily with surprising results. A rather tall object is viewed first standing and then lying down (Illustration 6c). When lying down the same object will seem to be less tall than before. The reason is that objects in the same plane are more easily defined and estimated than those above or below the observer.

COLOR | white light: the most highly colored light

White light is composed of the various colors of the rainbow. This fact can be observed and verified by using the glass pendant of a chandelier, the corner of a glass ashtray, or a prism. White light enters the glass and is dispersed; that is, the rays that make up white light will leave the glass at different angles. If these rays are projected onto a screen at a distance, the band of colors known as the spectrum will become visible.

The spectroscope, a scientific instrument designed to demonstrate this separation, will give a much better view of the colors of which white light is composed than will the edge of a glass ashtray. An eyepiece, similar to a telescope's, is used to observe the colors being analyzed. Spectroscopy is concerned with the measurement and determination of wavelengths in the spectrum.

1

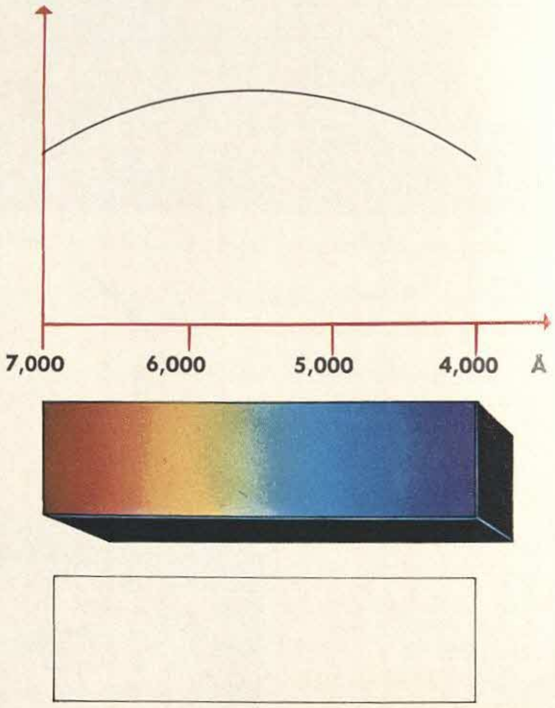
THE ANALYSIS OF WHITE LIGHT—White light can be classified into three different categories: hot light, such as the light of a tungsten filament lamp; cold light, such as the light of a mercury vapor lamp; and normal light, such as the light of the sun. The explanation of this classification can be seen by an analysis of the light of the sun.

The spectroscope reveals in sunlight the whole range of spectral colors from red to violet. Yellow immediately stands out as the brightest and most dazzling of all the colors. Red and violet, at the extremes of the spectrum, appear to be much weaker; although they are bright, they do not dazzle. Red and violet are almost equally bright, even though it is difficult to establish the relative brightness of two completely different colors. Yellow appears to be brighter for two reasons: first, the light of the sun contains more yellow than any other color; and second, the human eye is more sensitive to yellow than to other colors.

A graph showing the relative amounts of the various colors in the spectrum reveals these facts in a more precise way. The percentage of the colors present in the light of the sun is plotted along the ordinate, while the scale of the visible wavelengths is plotted along the abscissa. The wavelengths range from about 7,000 Å (Ångstroms) to about 4,000 Å (one Ångstrom is equivalent to 1×10^{-10} meters). 7,000 Å corresponds to dark red; 4,000 Å, to violet. The curve of the relative amounts of the different colors in the spectrum of white light is concave and reaches a maximum at the wavelength that corresponds to yellow.

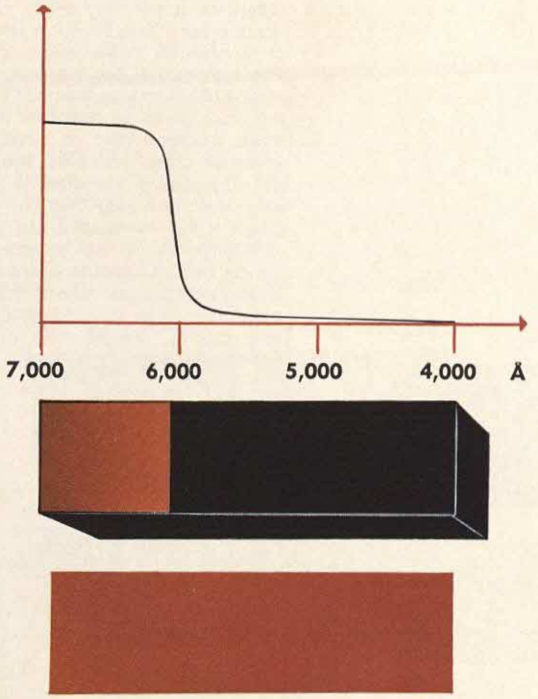
Sunlight is said to be normal light because it contains balanced amounts of all the colors of the spectrum, as can be seen from the accompanying graph. The light emitted from a tungsten filament lamp contains more red than yellow or violet; the light of a mercury vapor lamp contains more violet than red or yellow. A tungsten filament lamp, richer in red than sunlight, is classified as hot light because red is normally called a hot color. Lights that contain more violet tones are called cold.

In the experiment shown, the spectroscope was pointed directly at the sun. If it had been pointed at a sheet of white paper, or at the rectangle drawn below the spectrum in the diagram, the same result would have been obtained. This means that an object called "white" is one that reflects all the colors that it receives. Sunlight striking a sheet of paper still contains all the colors of the spectrum after it is reflected. The colors may be attenuated, but they are all weakened in the same manner, and the relationship between the intensities of the various colors remains unaltered. Thus, white light is the most colored of all light.

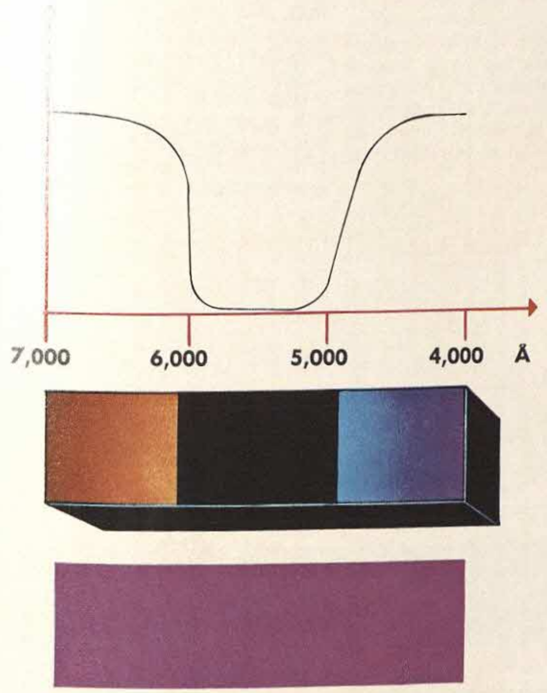


2

RED—Red is spectrally analyzed with the help of the colored rectangle in this illustration. If the rectangle is illuminated with a white light (for example, sunlight) only a part of the spectrum is seen in the spectroscope. The part seen is the part at the red end; the rest of the spectrum appears black. The yellow, green, blue, and violet bands are missing. These last colors, all of which are present in the white light used to illuminate the rectangle, have been absorbed by the colored paper. A red object is, therefore, said to retain the colors yellow, green, blue, and violet. These colors have been absorbed by the object and disappear within it. The color that is seen is the one reflected. A graph of the light reflected by a red body can now be drawn. It shows a horizontal line drawn above the abscissa between 7,000 Å and 6,000 Å. Between 6,000 Å and 4,000 Å the curve almost coincides with the abscissa; that is, the red rectangle reflects some small quantity of the other colors, but these amounts are so small that they are not perceived by the eye.

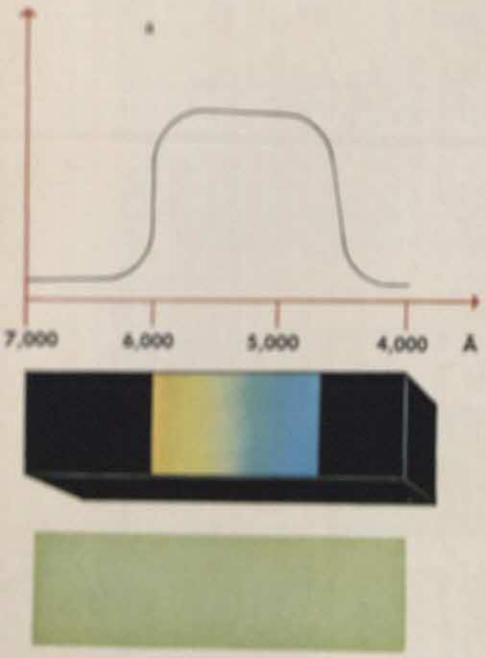


3

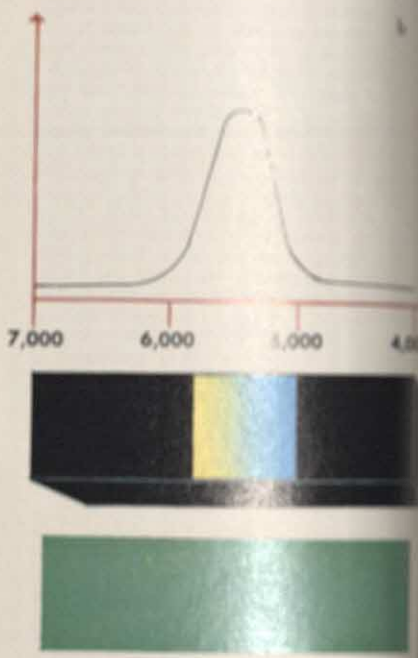


PURPLE OR MAGENTA—The light reflected in the spectroscope by this color is made up of two bands, with the yellow, green, and a part of the blue bands missing. The colors that remain in the light reflected by the rectangle are red, orange, dark blue, and violet. The graph for this color crosses the whole spectrum, but has a wide depression in the center of the graph.

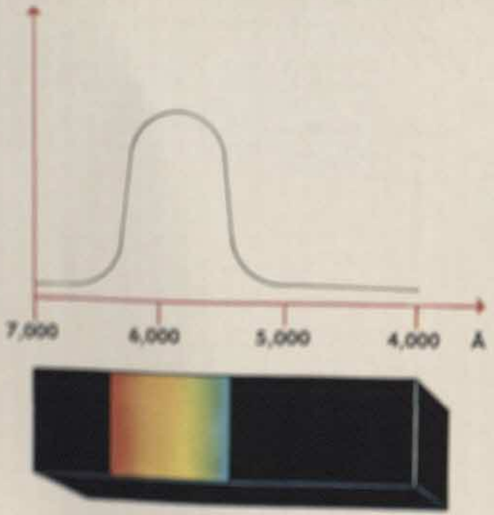
4



LIGHT GREEN AND DARK GREEN—The spectra of these two colors are different, but the difference is not very great. Light green reflects a band that covers a fairly wide range of wavelengths in the yellow and blue bands. The high part of the curve extends from almost 6,000 Å to less than 5,000 Å (Illustration 4a). Dark green, on the other hand, reflects a much smaller range of wavelengths in the yellowish green and blue bands (Illustration 4b). The peak of the spectral curve is, therefore, much narrower than the peak for light green. If the two spectra and the two curves are compared, it can be seen that spectral curves can be used to characterize a color. Because light green reflects more yellow than dark green, it is lighter than dark green and, consequently, reflects more of the spectrum than dark green.



5

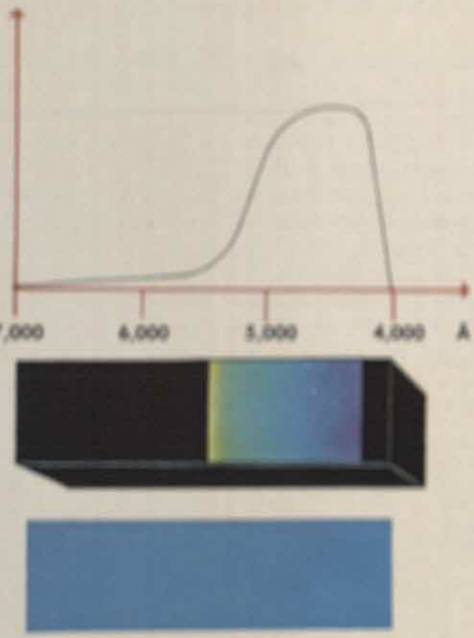


BRIGHT YELLOW—The spectroscope shows a broad spectral band for yellow. Orange, some red, yellow, green, and even a little blue comprise the color bright yellow. These colors, especially yellow and green, are the most luminous ones in the spectrum. Therefore, it is logical that the color that contains them will seem very bright. Another reason this color appears bright is that the human eye is particularly sensitive to radiations in these wavelengths.

6

BLUE—Blue reflects blue, a little green, and a little violet. Although it reflects a rather wide band of the spectrum, blue does not appear very bright, because the eye is not very sensitive to this section of the spectrum.

Nearly all the colors that have been examined have a very simple spectrum consisting of a single band. However, many colors have a more complex composition. Magenta, with its two distinct bands, is an example.

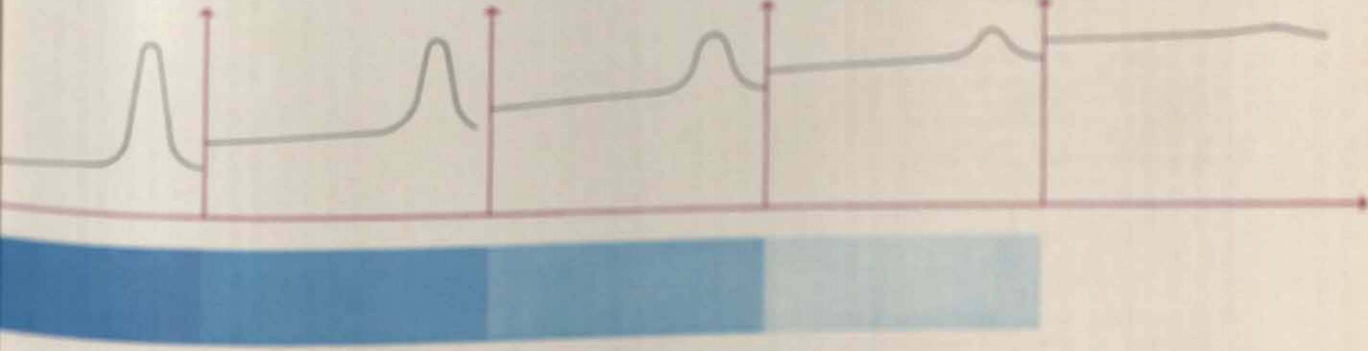


7

SATURATION—The term saturation may be best explained by this illustration, which shows strips of five different shades of blue, from dark blue on the left to light blue on the right. The spectral composition of these five blues is the same; the only difference is the amount of white that is mixed with the blue in each color.

This difference is easily observed with the spectroscope. If the dark blue rectangle on the left is examined with the spectroscope, its spectrum will be almost identical to the spectrum of the blue in illustration 6. Spectroscopic analysis of each of the successively lighter rectangles will reveal an increasing amount of all the other wavelengths of the spectrum. The spectral curve of the other rectangles will, therefore, approach the spectral curve of white, as the amount of white in each rectangle increases. The peak in the blue section of the spectrum will still remain, but it will become less marked as the amount of white in each color increases.

A color whose spectrum appears in the spectroscope as an isolated band, one that contains only minimal quantities of other colors, is said to be saturated. In this illustration the dark blue color in the rectangle on the left is saturated.



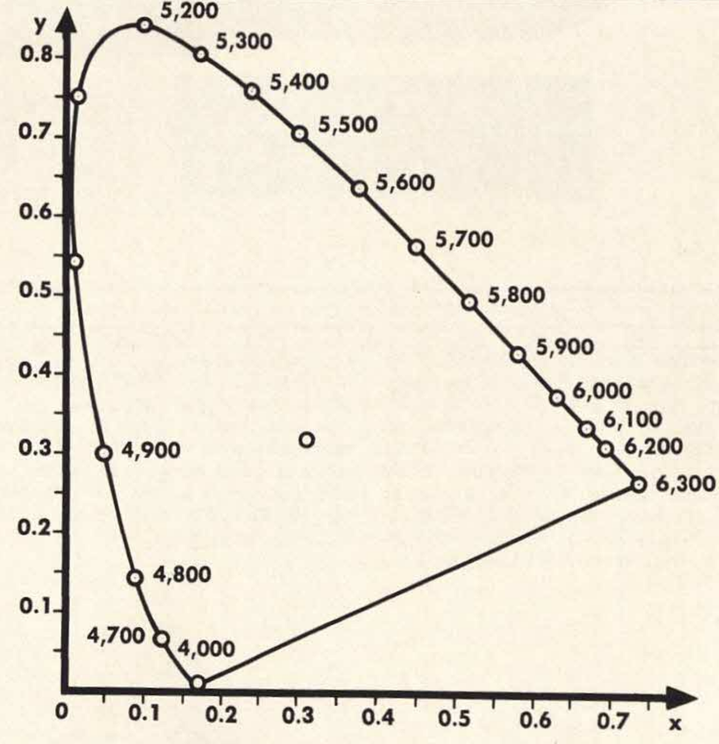
THE CHROMATICITY DIAGRAM

a graphic method of representing colors

Colorimeters are instruments used in evaluating the color of a surface under heterogeneous illuminations. In a typical colorimeter, half the field of view is occupied by the sample, which is illuminated in the desired manner. The other half is a comparison field whose color can be altered by the observer until a color match with the sample is established. In a colorimeter of the trichromatic type, the comparison field contains a mixture of red light, green light, and blue light, the amounts of which can be varied independently. Inasmuch as only one setting for each of the three controls will produce a match with a given sample, the color of the sample is indicated by the amounts of the three primaries used in making the match. Such primaries are often called basic stimuli, and the color of the sample is, therefore, said to be specified by its tristimulus values. (For a further discussion of tristimulus values, see the article "The Theory of Color," beginning on page 54 of this volume.)

The chromaticity diagram discussed here is a graphic method of representing the quality or chromaticity of color.

1



THE CHROMATICITY DIAGRAM—Ordinarily the quality of color is specified by three chromatic coefficients, which are defined in terms of the color's tristimulus values by the following equations:

$$x = \frac{X}{X + Y + Z},$$

$$y = \frac{Y}{X + Y + Z},$$

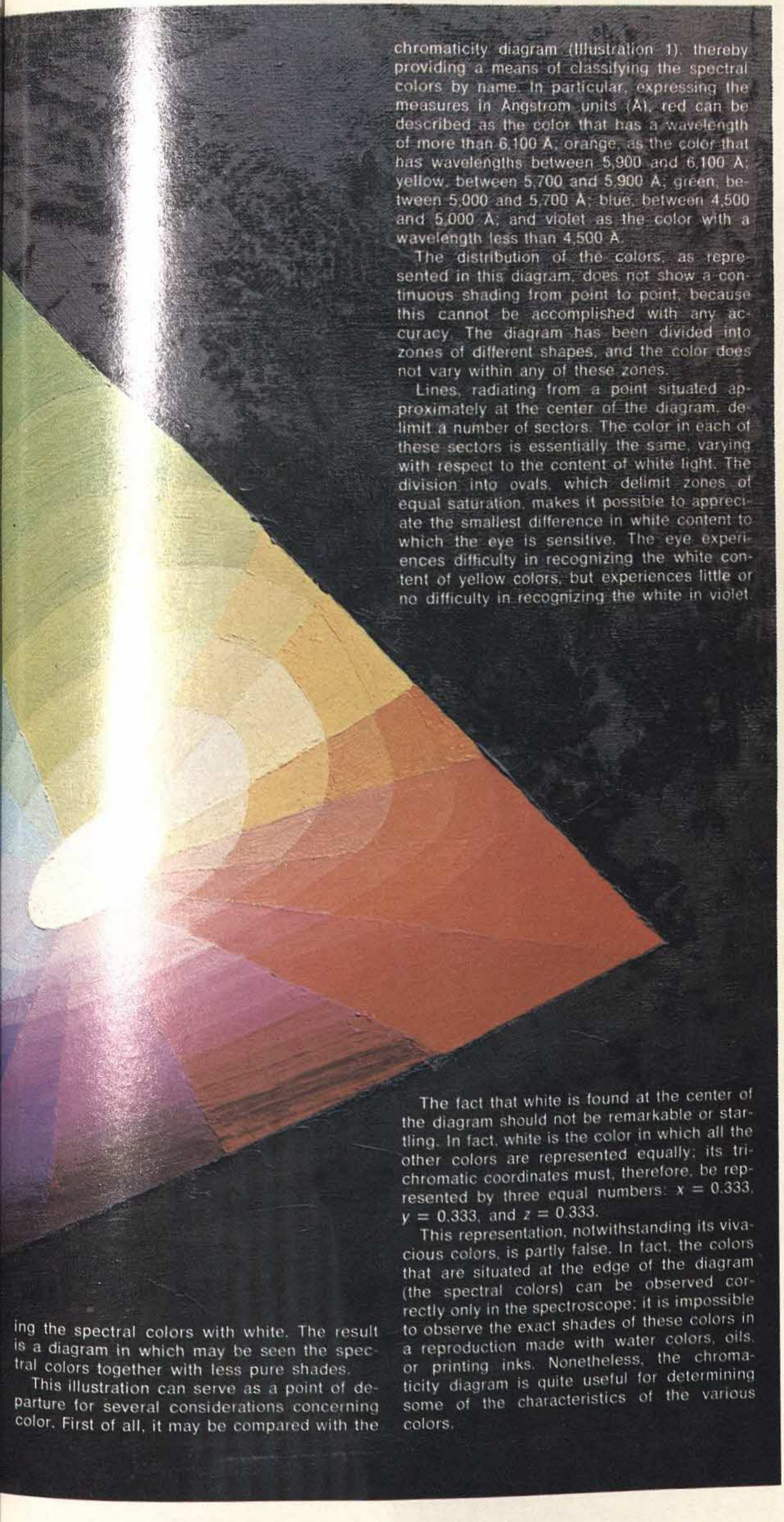
$$z = \frac{Z}{X + Y + Z}.$$

In other words, each color of the spectrum has a triad of numbers that may be used to define it. The three coefficients x, y, and z constitute the trichromatic coordinates or trichromatic coefficients. The sum of these coefficients is always equal to 1, thus obviating the necessity of specifying all three values. Usually, only the first two coefficients, x and y, are given for any color, because the third coefficient, z, can be obtained by subtracting the sum of x and y from 1. Consequently, the quality of any color can be plotted on a two-dimensional graph or diagram, using the values of x and y only.

If the values of x, y, and z are calculated or measured for all wavelengths of the spectrum, and the values of x and y are plotted, the curve shown in this illustration is obtained. The curve has the form of an arc; the wavelengths corresponding to the various points have been indicated. The form of this curve depends on the choice of the three fundamental colors in terms of which the coordinates of the others are expressed. This curve is called the locus of the spectrum and is drawn in the chromaticity diagram. This is a diagram plotted in an ordinary two-dimensional coordinate system in which both the abscissa and the ordinate are subdivided into fractions of unity. In fact, the locus of the spectrum is completely contained within a square having a side of length 1.

2

THE DISTRIBUTION OF THE COLORS IN TH
CHROMATICITY DIAGRAM—This illustrati
shows the approximate distribution of t
colors in the chromaticity diagram. The d
gram shows both the spectral colors and
the other colors that can be obtained by m



ing the spectral colors with white. The result is a diagram in which may be seen the spectral colors together with less pure shades.

This illustration can serve as a point of departure for several considerations concerning color. First of all, it may be compared with the chromaticity diagram (Illustration 1), thereby providing a means of classifying the spectral colors by name. In particular, expressing the measures in Angstrom units (Å), red can be described as the color that has a wavelength of more than 6,100 Å; orange, as the color that has wavelengths between 5,900 and 6,100 Å; yellow, between 5,700 and 5,900 Å; green, between 5,000 and 5,700 Å; blue, between 4,500 and 5,000 Å; and violet as the color with a wavelength less than 4,500 Å.

The distribution of the colors, as represented in this diagram, does not show a continuous shading from point to point, because this cannot be accomplished with any accuracy. The diagram has been divided into zones of different shapes, and the color does not vary within any of these zones.

Lines, radiating from a point situated approximately at the center of the diagram, delimit a number of sectors. The color in each of these sectors is essentially the same, varying with respect to the content of white light. The division into ovals, which delimit zones of equal saturation, makes it possible to appreciate the smallest difference in white content to which the eye is sensitive. The eye experiences difficulty in recognizing the white content of yellow colors, but experiences little or no difficulty in recognizing the white in violet.

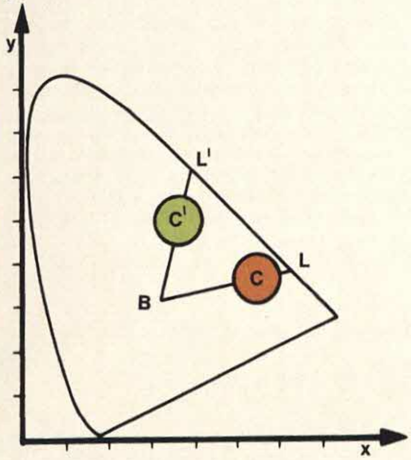
The fact that white is found at the center of the diagram should not be remarkable or startling. In fact, white is the color in which all the other colors are represented equally; its trichromatic coordinates must, therefore, be represented by three equal numbers: $x = 0.333$, $y = 0.333$, and $z = 0.333$.

This representation, notwithstanding its vivacious colors, is partly false. In fact, the colors that are situated at the edge of the diagram (the spectral colors) can be observed correctly only in the spectroscope; it is impossible to observe the exact shades of these colors in a reproduction made with water colors, oils, or printing inks. Nonetheless, the chromaticity diagram is quite useful for determining some of the characteristics of the various colors.

PURITY AND DOMINANT WAVELENGTH—The chromaticity diagram permits a rather immediate determination of some characteristics of a color. For example, in this diagram, two colors, an orange **C** and a green **C′**, are represented as points. Because the two colors in question are not perfectly pure, they do not occupy positions at the edge of the diagram. To determine the wavelengths of these colors, the points representing them must be joined to the geometric center of the diagram, the white point indicated by the letter **B,** and then these lines must be extended backward until they meet the edge of the diagram. At these intersection points can be read the highest spectral wavelengths of the colors under consideration. In this particular case, the values amount to about 5,500 Å for the green and about 5,950 Å for the orange.

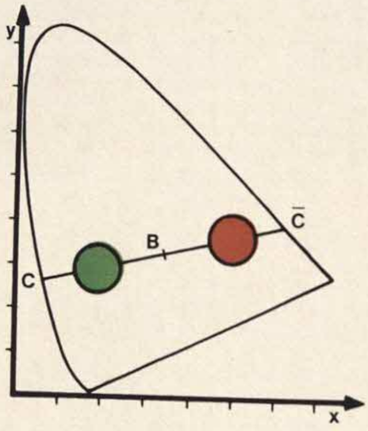
If the position of color **C** is plotted in the chromaticity diagram, the purity of the color can be defined as the ratio **BC/BL** multiplied by 100 (because purity is expressed as a percentage). In other words, the closer a color lies to the edge of the chromaticity diagram, the purer it is. If it lies near the edge, its purity is close to 100 percent. In the example, the purity of **C** is on the order of 75 percent, and the purity of **C′** amounts to 60 percent. **C,** therefore, is purer than **C′.**

3



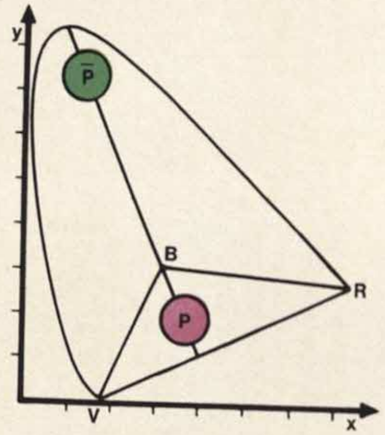
COMPLEMENTARY COLORS—The chromaticity diagram and the locus of the spectral colors delimiting it permit the formulation of a rigorous definition of complementary colors. If a point **C** is marked on the locus of the spectral colors, the complementary color can be found by joining the point **C** to the white center point **B** and then extending this line until it meets the other edge of the diagram. The intersection point represents the complementary color of **C.** The colors found in this particular case are shown in the two circles: a yellowish orange is the complementary color of a greenish blue.

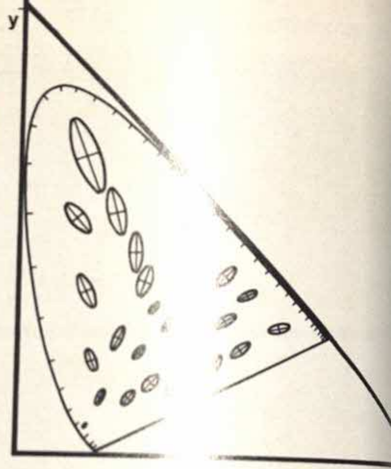
4



NONSPECTRAL COLORS—The locus of the spectral colors contains a tract, here represented by line **VR,** against which no wavelengths of spectral colors have been written. Colors represented by points in the triangle **VBR** cannot be specified directly by spectral wavelengths, but these colors can be specified indirectly by the wavelengths of their complementary colors. For example, if **P** is joined to **B** and the connecting line is extended beyond **B** to intersect the locus of the spectral colors, it reaches a point $\overline{\mathbf{P}}$. $\overline{\mathbf{P}}$ falls in the green part of the spectrum and complements **P;** the two colors combined form white, showing that **P** has some chromatic components lacking in $\overline{\mathbf{P}}$. Therefore, **P** can be called a "minus green" and the wavelength of the green that **P** lacks can be specified.

5



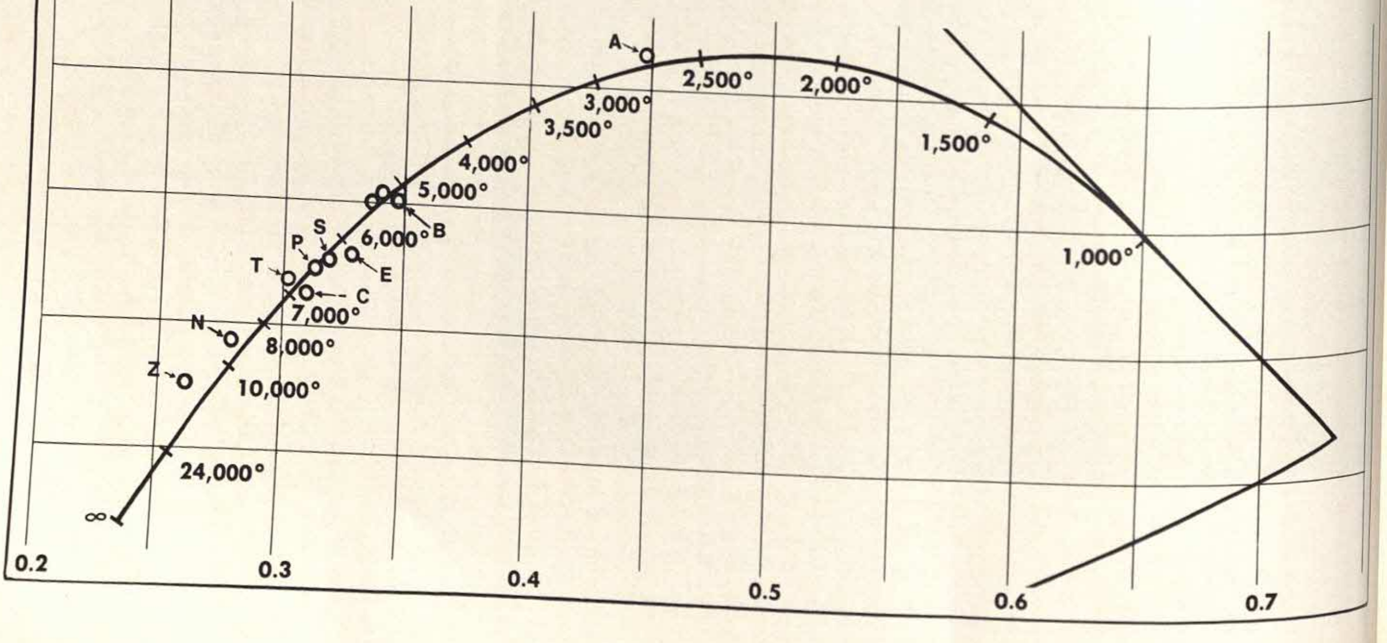


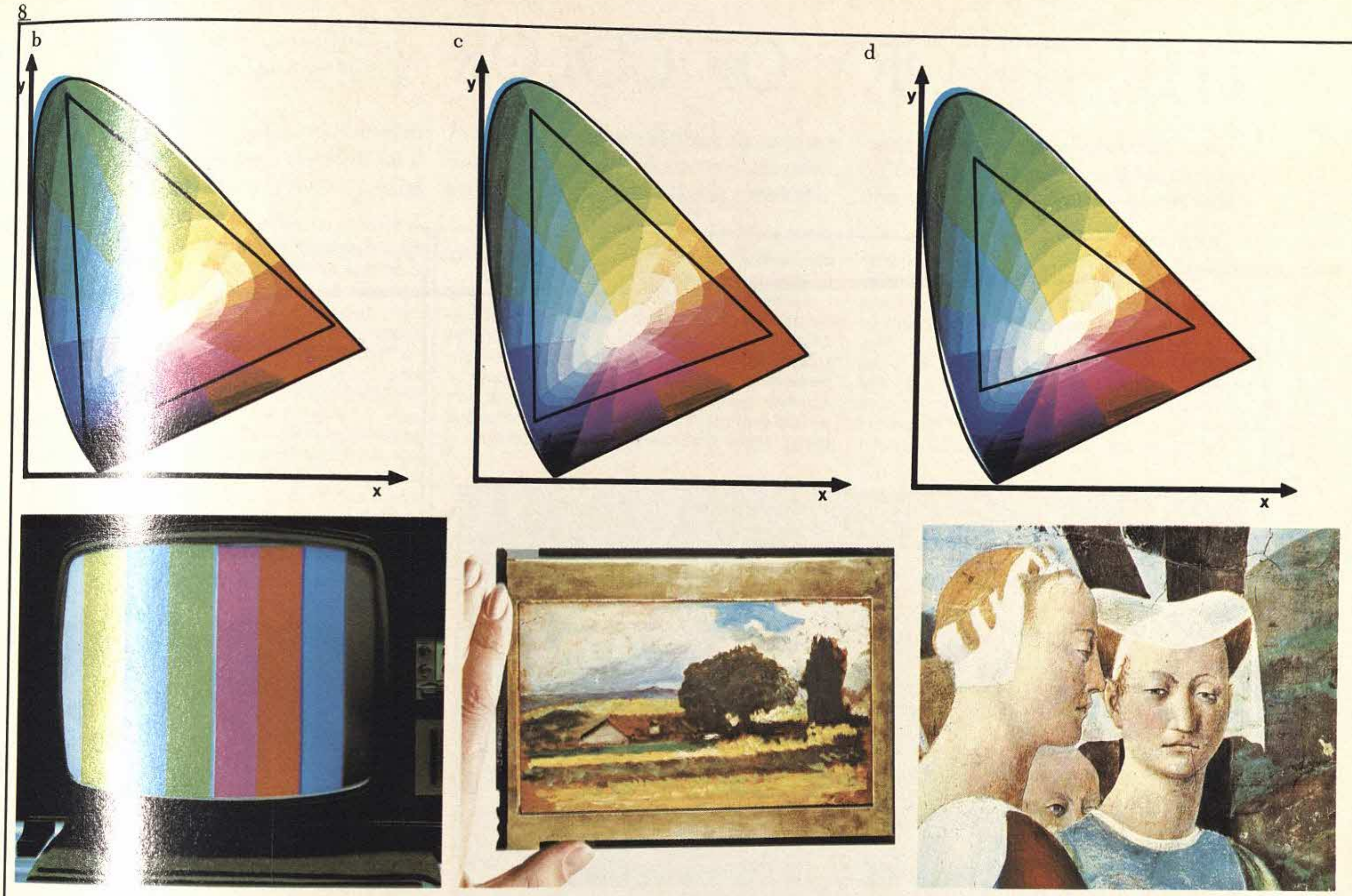
SENSITIVITY TO CHROMATIC DIFFERENCES—Some ellipses in this chromaticity diagram mark the areas within which the eye perceives practically the same color without being able to recognize the difference from one point to another. In fact, in illustration 2, the color remains uniform over a rather large area; the eye can easily perceive the difference between one area of uniform color and the adjacent area. If this diagram were reconstructed with areas so small that only a minimal difference of color existed between any two adjacent areas, the diagram would have a practically perfect and continuous variation of shades. In the diagram the area within which color differences (other than those between opposite points on the border of the area) are not recognizable is always represented by an ellipse whose size varies according to its position.

COMPLEMENTARY COLORS — The chroma-sured in terms of its coordinates and is plotted as a point in the chromaticity diagram, the point obtained is known as the color point. But what is the color point of "white," the name given to an extensive class of colors, as can be seen in this chromaticity diagram which plots the color points of many whites. For example, **E** is the white point indicated by the letter **B** in the previous diagrams; it represents the color of a hypothetical source that emits the same energy on all the wavelengths of the visible spectrum. A second fictitious source known as Illuminant **C** is defined by the International Commission of Illuminations. The equality of the white light of source **C** approaches that of sunlight, while the source known as Illuminant **A** approaches the light emitted by tungsten filament lamps.

Other diagramed illuminants often taken as points of reference for white light include **S,** the white point of sunlight when the sky is perfectly clear. The background light of the sky is blue, distorting the color of the light emitted by the sun. The white point of the source, consisting of the sun and the sky's blue, is represented by **T.** When the sky is covered by clouds, the light that filters through has its white point, represented by the letter **P.** With the exception of **A,** all these white points are quite close together; **A** is displaced toward the red. If, on the other hand, the light of the sky alone is considered as the illluminant, points may be found that are strongly displaced toward the violet part of the diagram. In fact, **N** is the white point that reproduced the white of the sky's light at the celestial north pole; **Z** is the white point of the sky's light at the zenith. Because the sky is blue, the corresponding white points are displaced toward the blue part of the spectral colors. Finally, the curve that crosses the diagram in an oblique direction—starting at the color points of the sky and arriving at the spectral red of the longest wavelength—represents the locus of white points of the radiation emitted by a black body at various temperatures.

7





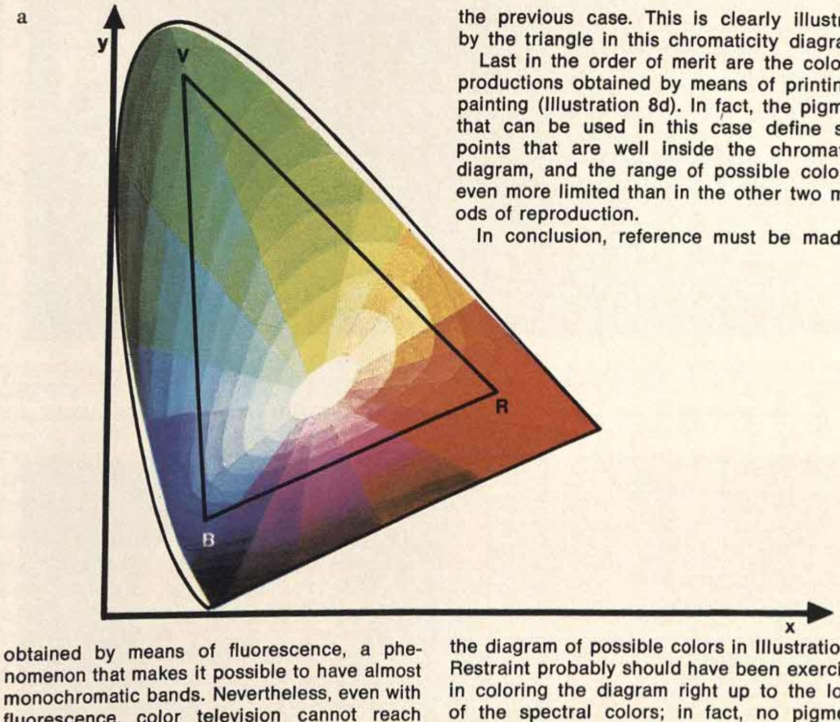
AN APPLICATION: THE QUALITY OF COLOR REPRODUCTION

—Examination of the chromaticity diagram aids in the evaluation of the quality of color prints or reproductions.

In Illustration 8a, three points, **B, V,** and **R,** are marked inside the locus of the spectral colors; these points correspond to the colors of three pigments or three colored radiations that are used for composing the colors to reproduce a figure. Real colors are involved here; therefore, these points lie not on the locus itself but somewhat inside it. The colors that can be formed by mixing suitable quantities of these three colors are all to be found within the triangle **BVR;** a choice of pigments always limits the range of colors that can be obtained. Even if it were possible to obtain three pigments whose color points lie on the locus of the spectral colors, a triangle would still be inscribed within the locus of the spectrum, inasmuch as this locus is everywhere curved and always outside a straight line joining any two points on it. Consequently, some rather large areas of fairly pure colors that cannot be reproduced would always remain.

In order to overcome this difficulty, the three fundamental colors of the International Commission on Illumination (Illuminants A, B, and C) have been chosen from among fictitious colors that lie outside the locus of the spectral colors. The triangle defined by these colors, therefore, contains the whole of the chromaticity diagram—that is, the entire range of possible colors. In fact, the coordinates of these three colors are (0,0), (0,1), and (1,0).

In practice, the systems for reproducing colored images are always imperfect. However, a scale of merit can be established. The place of honor in such a scale is occupied by color television (Illustration 8b). The colors that can be reproduced on a television screen can be better than those obtainable by means of any other system. In fact, these colors are obtained by means of fluorescence, a phenomenon that makes it possible to have almost monochromatic bands. Nevertheless, even with fluorescence, color television cannot reach the color points lying on the locus of the spectral points. In this diagram the positions of three possible color points of a color television system are shown.



The system of color reproduction next in order of merit is the one that can be obtained by means of the projection of slides (Illustration 8c). In this case a luminous image is projected on a screen, but the range of possible colors is rather more restricted than it was in the previous case. This is clearly illustrated by the triangle in this chromaticity diagram.

Last in the order of merit are the color reproductions obtained by means of printing or painting (Illustration 8d). In fact, the pigments that can be used in this case define some points that are well inside the chromaticity diagram, and the range of possible colors is even more limited than in the other two methods of reproduction.

In conclusion, reference must be made to the diagram of possible colors in Illustration 2. Restraint probably should have been exercised in coloring the diagram right up to the locus of the spectral colors; in fact, no pigments suitable for coloring the areas close to this locus can be found; and, even if such colors existed, they could not be reproduced with the perfection required by the theory. Only through spectroscopic examination of the colored lights in the spectrum of white light is it possible to view the purest colors that can be perceived by the eye. For this reason, the sight of such beautiful colors is always fascinating —even though they are seen frequently.

THE THEORY OF COLOR

infinite combinations of fundamental colors

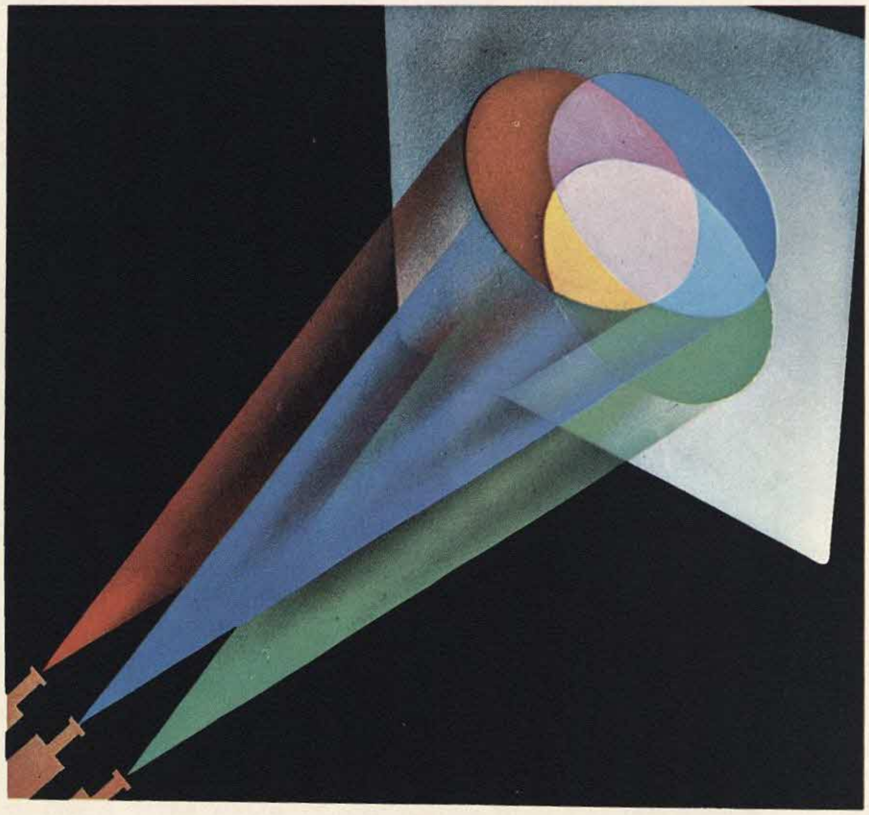
One of the most interesting characteristics of the human eye is the fact that it is able to react to the wavelength composition of the different radiations that strike it.

In other words, the eye is capable of different sensations according to whether it is struck by electromagnetic light radiations of short, medium, or long wave-

ADDITIVE COMBINATION — This illustration shows the meaning of the additive combination of three radiations. Each of three projectors throws a beam of light onto a white screen on which the beams form three colored circles. The three circles are so arranged as to be partially superposed, and a radiation that comes close to being white occurs in the area common to all three; nevertheless, the color of this area is not a perfect white. (The color can be made to approach white more closely by varying the intensity of the beams.)

A single color formed as the sum of several colors is a phenomenon called synthesis. Because the three radiations are "added" on the screen, the phenomenon is an additive synthesis or an additive process. This concept of additive synthesis is used to establish the principle of color measurement and to define the so-called tristimulus values of a color.

1

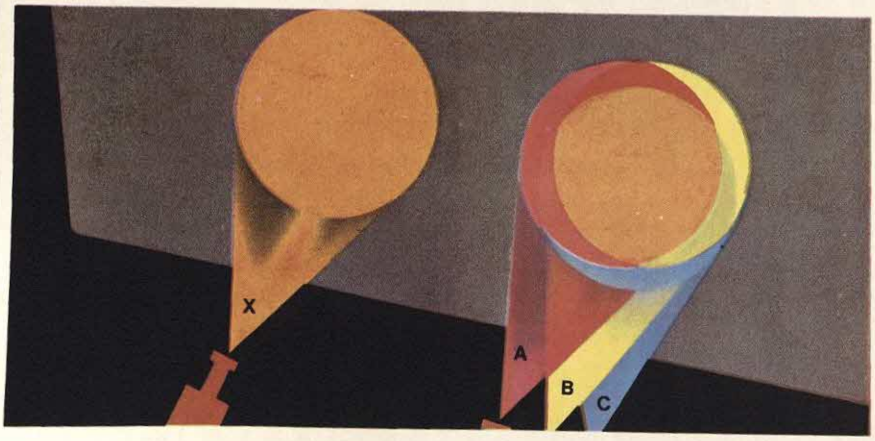


MEASURING A COLOR BY MEANS OF THREE COMPONENTS—With three arbitrarily chosen colors it is possible to obtain a triad of numbers that can serve to give a measure of some related color.

The illustration shows the three colors chosen as fundamental colors for measuring the others. These colors are **A,** blue, **B,** yellow, and **C,** violet. In whatever proportions these three colors are mixed in additive synthesis, the color **X** cannot be duplicated. Even so, a method can be devised that will enable measurement of **X** in terms of **A, B,** and **C.** If **X** could be composed as a mixture of a suitable percentage of each of these three colors, units of measurement characteristic of **A, B** and **C** could be chosen, and the three per-

3

2

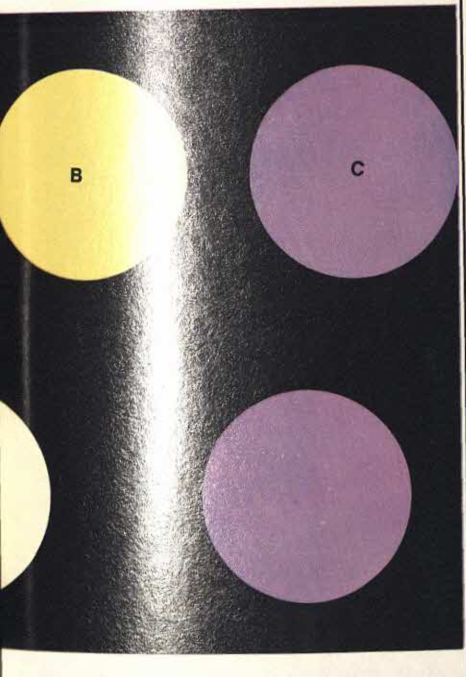


COLOR COMPARISON BY SYNTHESIS—This illustration shows the principle on which, however indirectly, the measurement of color is based. A projector on the left throws a beam of light onto the screen; the color of this light may be called **X.** To find a measure for this color, three other projectors are set up; each emits light that has passed through a different filter. The three projectors are so arranged that their beams will strike the screen at the same point. Each of the three projectors is fitted with a device that enables the intensity of the beam of light to be varied. Thus, if the three colors are **A, B,** and **C,** any desired quantity of **A** may be added to any desired quantity of either **B** or **C,** or of both.

At this point, nothing has been said about the criterion that governed the choice of the three colors **A, B,** and **C** that are to be combined so that their sum will be equal to **X.** At length. The sensation that corresponds to the perception of these different wavelengths is the one designated briefly by the word *color*. The word indicates something that is closer to the psychology of sensations than it is to physics, and it can express two completely different things. For example, consider the two sentences: "The traffic light is green at this moment"

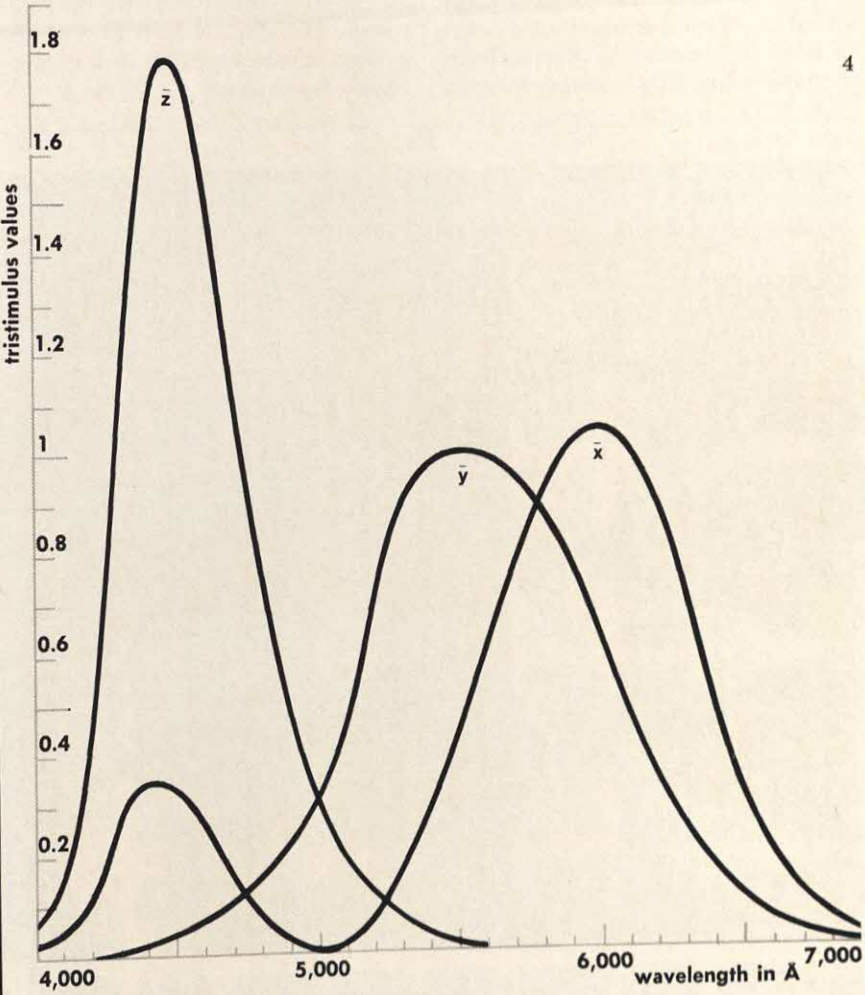
centages would then provide a measure of **X.** Because this is not possible, however, another method must be found that has general validity.

To begin, it is not possible to compose **X** by mixing **A, B,** and **C,** but it is possible to match a mixture of **X** and **A,** for example, by suitably combining **B** and **C.** In this particular case the additive mixture of **X** and **A** yields a purple color, and the same color can also be obtained by mixing a little of **B** with a lot of **C.** Therefore, a measure of **X** can be given by saying how much of color **A** (with a negative sign) must be added to **X** in order to make it equal to the color obtained by mixing **B** and **C** in a certain proportion. This principle, subject to a suitable choice of the three colors, enables measurement of any color whatsoever.



first it may seem possible to choose any three colors; for example, yellow, green, and blue. By suitably treating these three colors an infinity of intermediate colors can be obtained; thus, an equally large number of colors such as the color **X** can be matched. However, no matter how versatile this method may seem, it cannot be used to compose all the colors that exist in nature.

A first extension of the method consists of choosing the three colors **A, B,** and **C** (illustrated) in such a way that they will be evenly distributed in the spectrum of white light, rather than being concentrated around the green part of the spectrum as was the case with the three colors used in the first example. Although this wider range of basic colors allows an even larger number of colors to be produced than with the first three colors chosen, it still does not produce all colors.



4

TRISTIMULUS VALUES OF SPECTRAL COLORS—The definition of the three fundamental colors that serve to measure all others was prepared in 1931 by the International Commission of Illumination (ICI). These three colors do not really exist; they are "fictitious" colors. By mixing them in appropriate proportions, it is possible to obtain any real color.

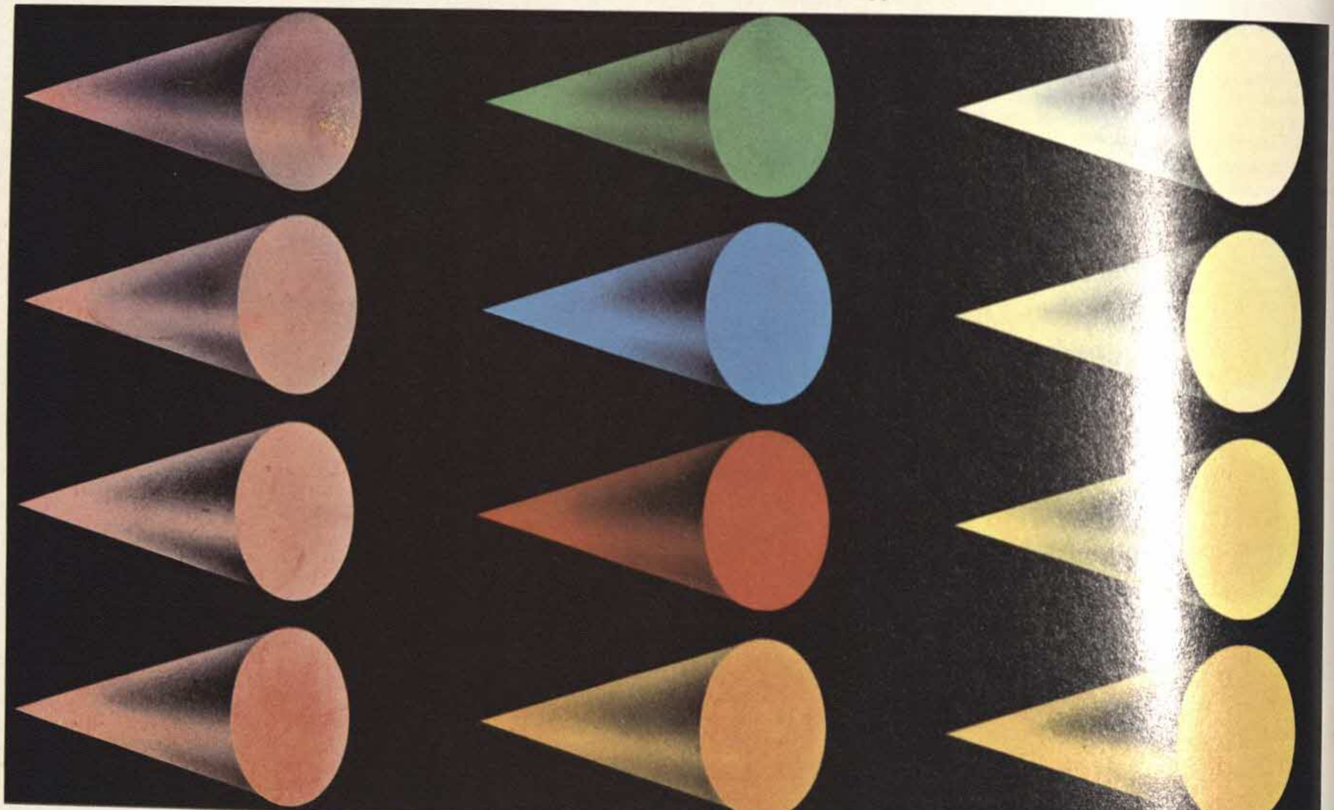
Various methods are used to define these three fundamental colors. One of these methods (illustrated) consists of graphing three curves that represent the quantities of the three colors needed to compose all the colors of the spectrum. The abscissa of the graph, therefore, shows the values of the wavelengths of the visible spectrum; the ordinate shows values that are proportional to the composition of the mixture. For example, the light of 5,000 Å can spectrum having a wavelength of 5,000 Å can be produced by means of 0.0049 of component $\bar{x}$ (magenta), 0.3230 of component $\bar{y}$ (green), and 0.2720 of component $\bar{z}$ (green). These three values are called the tristimulus values of the color chosen. By means of the curves in this graph, the tristimulus values of all the other colors of the spectrum can be found. The three curves are called the curves of the ICI standard colors. They are represented by means of the symbols $\bar{x}$, $\bar{y}$, and $\bar{z}$.

It should be kept in mind that whenever the synthesis or sum of two colors is mentioned, additive synthesis is inferred. The synthesis of two commonplace colors is obtained by mixing paints or pigments; this is the subtractive process of color combination. The color obtained by mixing two pigments is not the same as that obtained by mixing two light rays that have exactly the same two colors. The additive synthesis of a red and a green, for example, yields a color that is quite close to white; the mixture of a red and a green pigment, however, yields a brown color. In order to predict the color that will be obtained as a result of the additive synthesis of two colors, the spectrum of the one must be added to the spectrum of the other; the examination of the combined spectrum then enables deduction of the new color.

and "The cover of this book is green." In the first sentence the term "green" indicates the quality of a beam of electromagnetic radiation; that is, the composition of the beam in terms of wavelength. In the second sentence the term indicates the property of a body—the property of altering the characteristics of the light that strikes it so as to transform white light into green. In both cases, however, the physicist requires a method of measuring color so that this term has a precise physical significance rather than a purely subjective meaning.

The part of physics that deals with the problem of measuring color is called the theory of color; it is not widely known, even though it is the object of studies on the part of physicists.

5a 5b 5c



THREE CHARACTERISTICS OF LIGHT RADIATIONS—As an introduction to the study of color, three important characteristics of light rays must be considered. These are intensity, predominant wavelength, and purity. The intensity of light can be defined as an energetic property of the beam; the remaining two magnitudes, on the other hand, are chromatic characteristics.

Illustration 5a shows four beams of radiation falling on a screen that has equal reflecting power for light of every wavelength; the reflecting power of the screen is extremely high; the screen is said to be white. It reflects or diffuses the light radiation without altering its color composition. The four light rays in Illustration 5a all have the same chromatic characteristics; all four are pinkish, yet each has a different intensity. A thermopile or a bolometer would clearly demonstrate that the four beams have different intensities (energies).

The four beams of colored light in Illustration 5b have the same intensity of energy, but the predominant wavelength is different in each case. The first is not a pure green, but no one would identify it as any other color. The second beam is blue, the third one reddish, and the fourth one yellowish. These colors are not the green, the blue, the red, and the yellow that can be seen in the spectrum of white light, but simply colors that are close to them. Nevertheless, because the spectral colors closest to the four colors shown are respectively known as green, blue, red, and yellow, the dominant radiation in each case has the color of the appropriate spectral radiation. Another fact that must be borne in mind is that the radiations of the four beams, by virtue of their different colors, appear to be of different intensities even though, from the point of view of energy each of the four beams has exactly the same intensity. The fact that a beam of blue light is perceived as being darker than a yellow beam is only due to the differential sensitivity of the human eye.

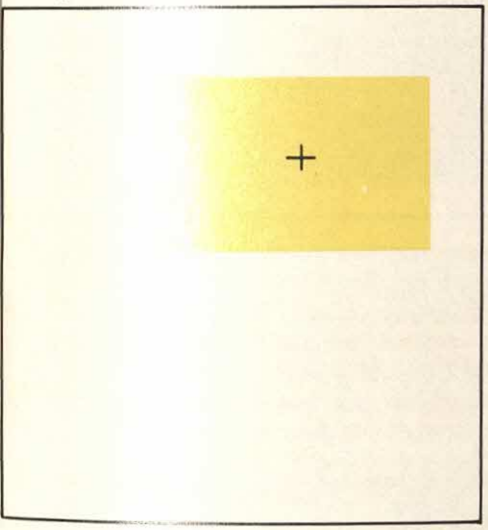
The four beams in Illustration 5c show the meaning of purity. The fourth beam is the purest of all: it is made up of a spectral yellow. The other three beams contain other colors; the first contains a considerable quantity of other colors, so much so that it appears almost white with a slight yellowish tinge. The others are gradually purer—containing decreasing quantities of radiations of other wavelengths. The more nearly a beam contains a single wavelength, the purer it will be. The purity of the beams increases from the first to the last.

THE APPEARANCE OF COLOR

illusions involving hues

Optical illusions involving color provide useful clues for the understanding of two distinct categories of phenomena affecting human vision: physiological phenomena, centered in the eye, where light energy is transformed into chemical energy for transmission over the optic nerves to the brain; and psychological phenomena, centered in the brain, which interpret or make judgments concerning the messages received. Color is a very important component of these messages; it not only discloses the physical nature of objects in the environment, but it influences human reaction to that information. For example, when a patient is subjected to a series of Rorschach tests, in which a large number of black-and-white pictures are followed by a sequence of colored pictures, it has been observed that the interpretations of the black-and-white pictures may have been unhesitating, but as soon as the first colored picture is shown, the patient becomes disoriented.

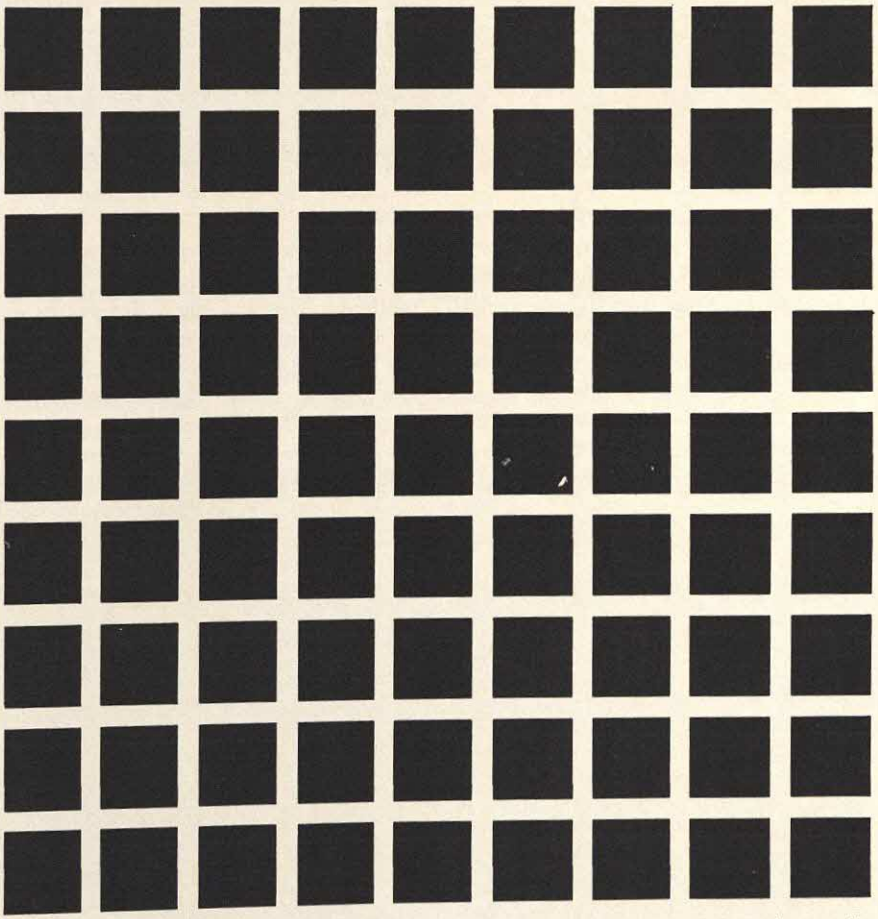
1



THE EFFECT OF FATIGUE—This illustration shows an area colored yellow, a hue of great spectral purity. If it is observed for a minute or so while well illuminated, the receptors in the retina sensitive to yellow will become fatigued. If the eyes then move to a completely white surface, an area of the same form will appear, but it will have a purple-violet hue, the color complementary to yellow. The eyes actually see the area on which it is concentrating, minus the very hue that has fatigued the yellow-sensitive receptors; this leaves, of course, the sensation of white minus yellow, or the complementary color of yellow. The fact that the purple-violet area is similar to that of the yellow area previously observed demonstrates that the phenomenon is physiological, not psychological—that is, it occurs before the image reaches the brain.

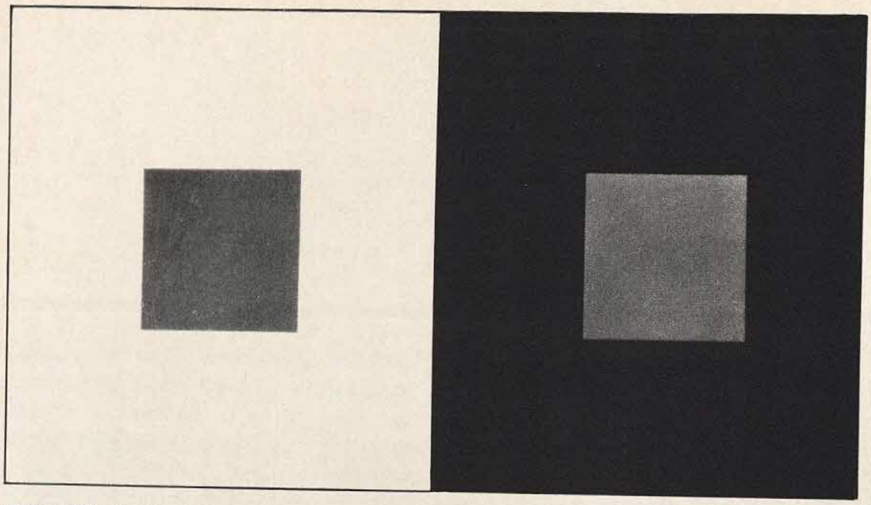
The theory already mentioned is not universally accepted as the explanation of the response of the retina to the stimuli of colored radiations. Another theory, which also postulates three kinds of sensitive receptors, suggests that the receptors are sensitive to pairs of colors: yellow-blue, green-red, and black-white.

2



THE EFFECT OF CONTRAST—If these black squares separated by white spaces are observed, gray areas will appear at the intersections, but only at those sections of the total area that are not at the center of vision. If the eyes are moved to concentrate at the intersections where the dark areas appear, they immediately disappear; and as soon as the eyes are moved from them, they immediately reappear. It should be noted that whereas, in the illusion of the afterimage in the preceding illustration, the complementary color only appeared after the original hue was under observation for a period of time, in this instance the illusory dark areas appear and disappear and reappear with no lapse of time other than that required to glance from one part of the illustration to the other.

3



INTENSITY—On the left side of this illustration, a gray square is surrounded by a white background; on the right side, a gray square of precisely the same intensity is surrounded by a black background. The gray area on the left, however, appears to be darker than that on the right. This results from a psychological interpretation in the brain, which insists on comparing or contrasting the smaller square with its background, the larger square.

In this and some of the following illustrations, it is possible to test the real against the apparent quality of the illusion by masking out the areas responsible for the illusion. Thus, a sheet of opaque paper in which an area the size of the smaller squares has been cut out can be set over both larger squares, masking them; it then becomes apparent that the two gray squares, set within an identical background, are indeed of the same intensity.

COMPLEMENTARY COLORS FOR EXPERIMENTS IN CONTRAST—If a given hue is set against two complementary colors at a high level of saturation, a strong contrast would be expected; but considerable effect can also be obtained with complementary colors of low saturation. Some spectral colors at a high level of saturation are shown in Illustration 5a. (Each vertical pair, in this and the following groups, represents a set of complementary colors.) The same colors, but at a low level of saturation, are shown in Illustration 5b. In Illustration 5c appear colors obtained by combining several pure spectral colors, resulting in hues of medium saturation, duller, and less lively.

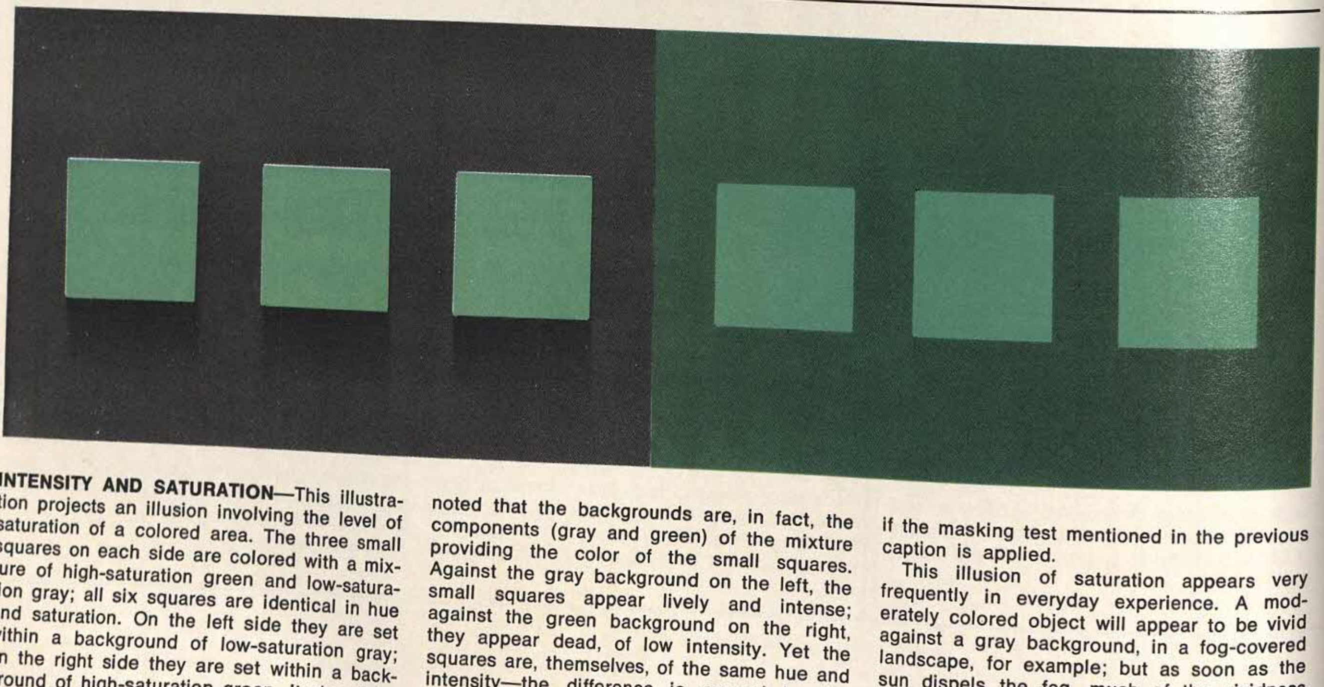
These colors are used, in the next illustration, for an experiment on the backgrounds of color. They are best reproduced by using a tempera coloration that will dry on an opaque surface, preferably pieces of cardboard. At the center of these pieces of colored cardboard can be placed a smaller piece of cardboard of a different color. As the smaller pieces are moved from one background sheet to another, the change in their appearance becomes evident. An even better, but more difficult method, is to apply the two complementary colors to two sides of a rectangular area and then color the center of the area with the color to be observed for contrast; this method requires a great deal of care so that the complementary colors will not merge with each other, but will be divided by a neat line. If this is done carefully, a more perfect contact results between the background colors and the color under observation.

This indicates the effect of color on the unconscious mind.

Some colors are considered to have a calming effect, whereas others tend to excite. Blues and bluish grays, described as cold tones, when used to cover a wall surface, seem to make a room larger; the warm tones ranging from red to yellow seem to make a room smaller. Some colors are easily perceived amid others; some remain relatively obscure. Such effects are exploited in advertising, in which the essential graphic objective is to attract attention to particular visual subjects.

In physiological terms, color is perceived by sensitive elements (receptors) on the retina of the eye. These receptors are divided into three categories, according to the hues to which they are sensitive—yellow, red, and bluish-green (cyan). White light stimulates all three kinds of receptors; red light stimulates only those sensitive to red; yellow only those sensitive to yellow; and cyan only those sensitive to cyan. If the receptors of any one category become fatigued as a result of overexposure to a particular hue, the eye perceives another color in an altered form because the fatigued receptors cannot perceive the hue to which they are ordinarily sensitive at normal intensity.

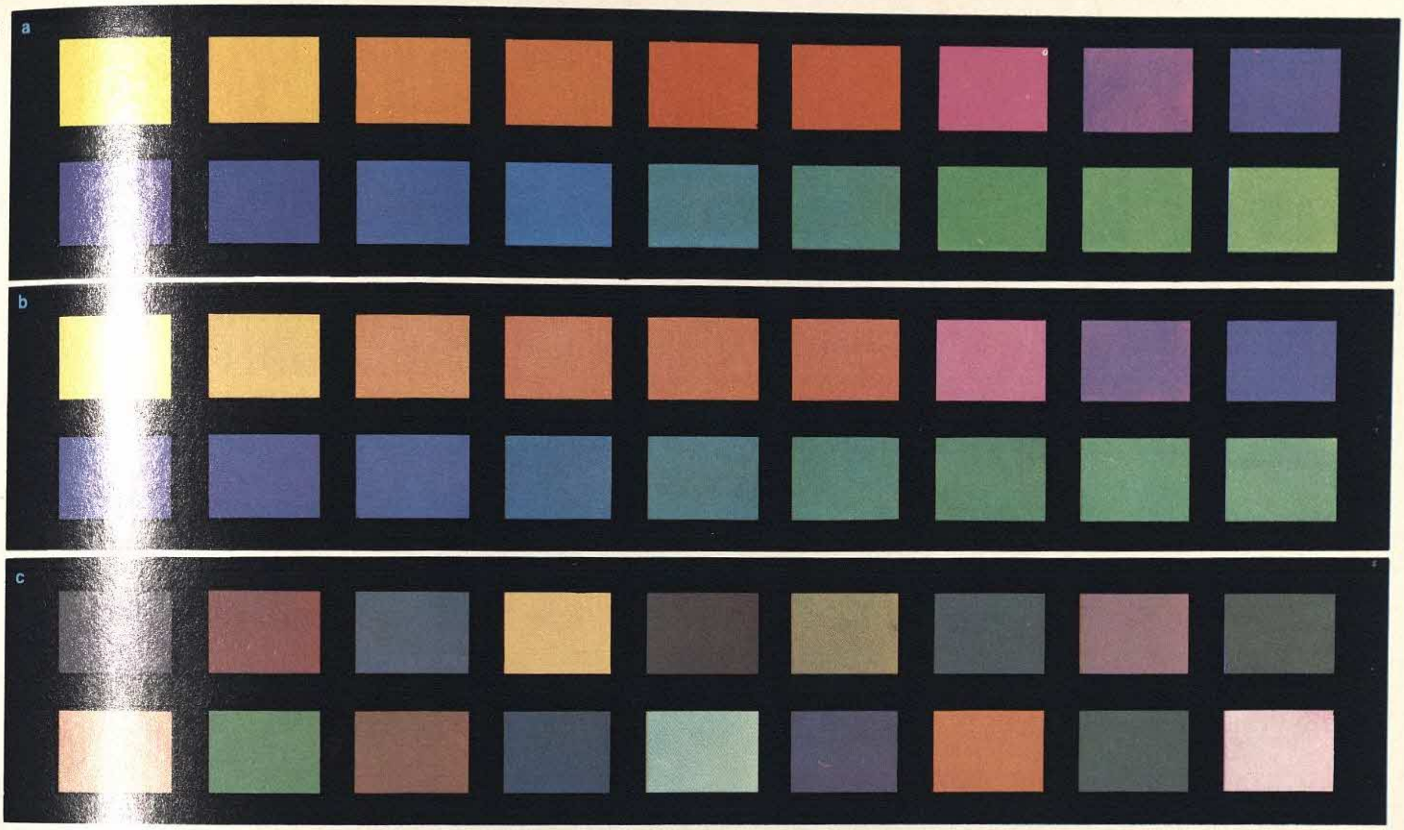
4



INTENSITY AND SATURATION—This illustration projects an illusion involving the level of saturation of a colored area. The three small squares on each side are colored with a mixture of high-saturation green and low-saturation gray; all six squares are identical in hue and saturation. On the left side they are set within a background of low-saturation gray; on the right side they are set within a background of high-saturation green. It should be noted that the backgrounds are, in fact, the components (gray and green) of the mixture providing the color of the small squares. Against the gray background on the left, the small squares appear lively and intense; against the green background on the right, they appear dead, of low intensity. Yet the squares are, themselves, of the same hue and intensity—the difference is caused by the contrast of the background, as will be evident if the masking test mentioned in the previous caption is applied.

This illusion of saturation appears very frequently in everyday experience. A moderately colored object will appear to be vivid against a gray background, in a fog-covered landscape, for example; but as soon as the sun dispels the fog, much of the vividness will vanish, while the countryside itself will be brighter.

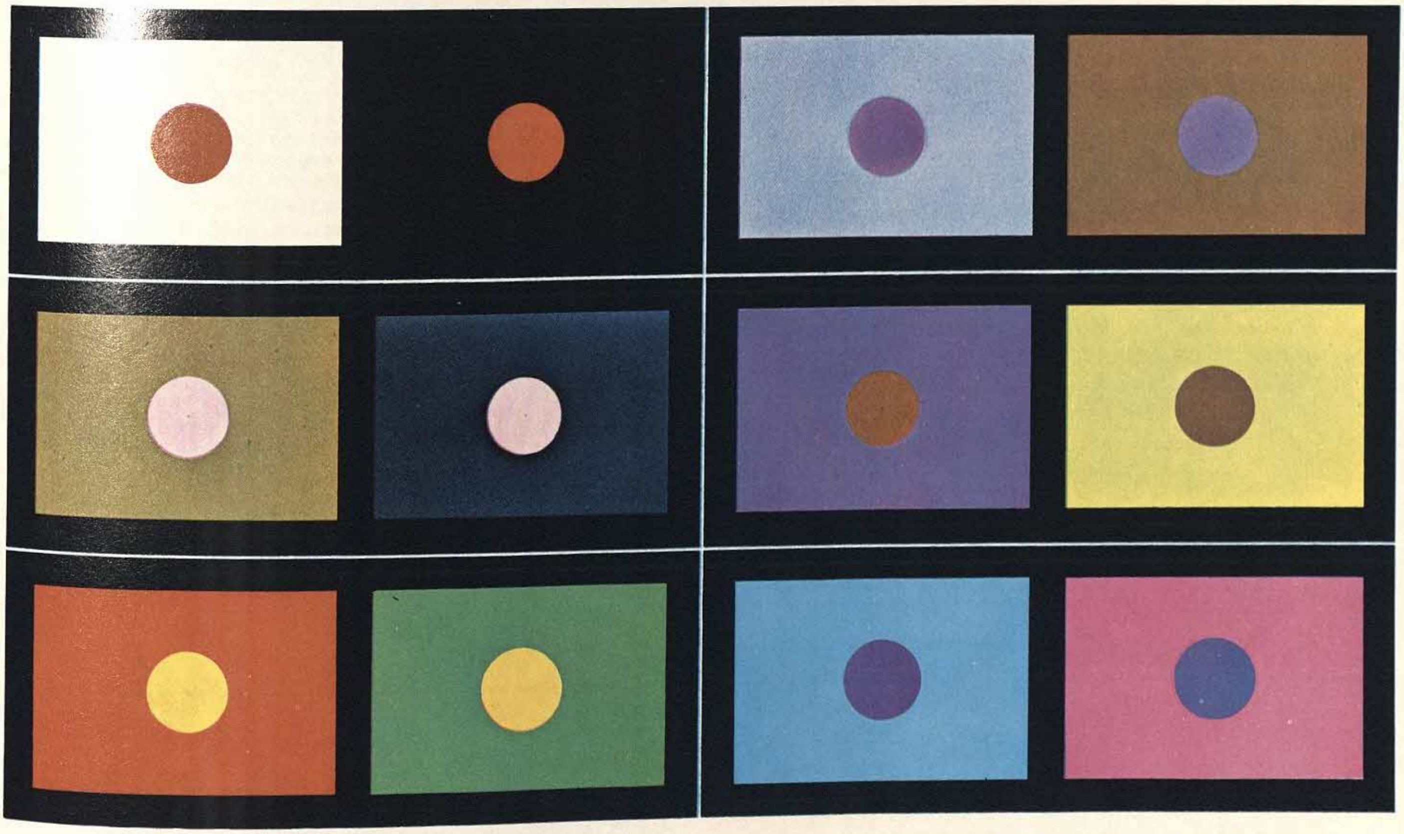
5



THE EFFECT OF BACKGROUND ON COLOR— This illustration shows how the same color appears to be different when viewed against different backgrounds. The background colors should be complementary, to provide the maximum difference in contrast with the color in the center, and the latter should also be somewhat different from each of the background colors. Once again the masking technique is useful to demonstrate that the central colors are identical no matter how different they appear before being masked.

Even the pure spectral colors such as red and yellow are very sensitive to the color of the background; however, pastel shades provide the greater contrast. The low-saturation purple in the upper right corner, for example, appears very different when seen against the brown background from the way it appears against the violet background in the panel to the left. In any case, pastel shades and mixtures of many colors are more sensitive to this effect and seem to differ markedly against different backgrounds.

6



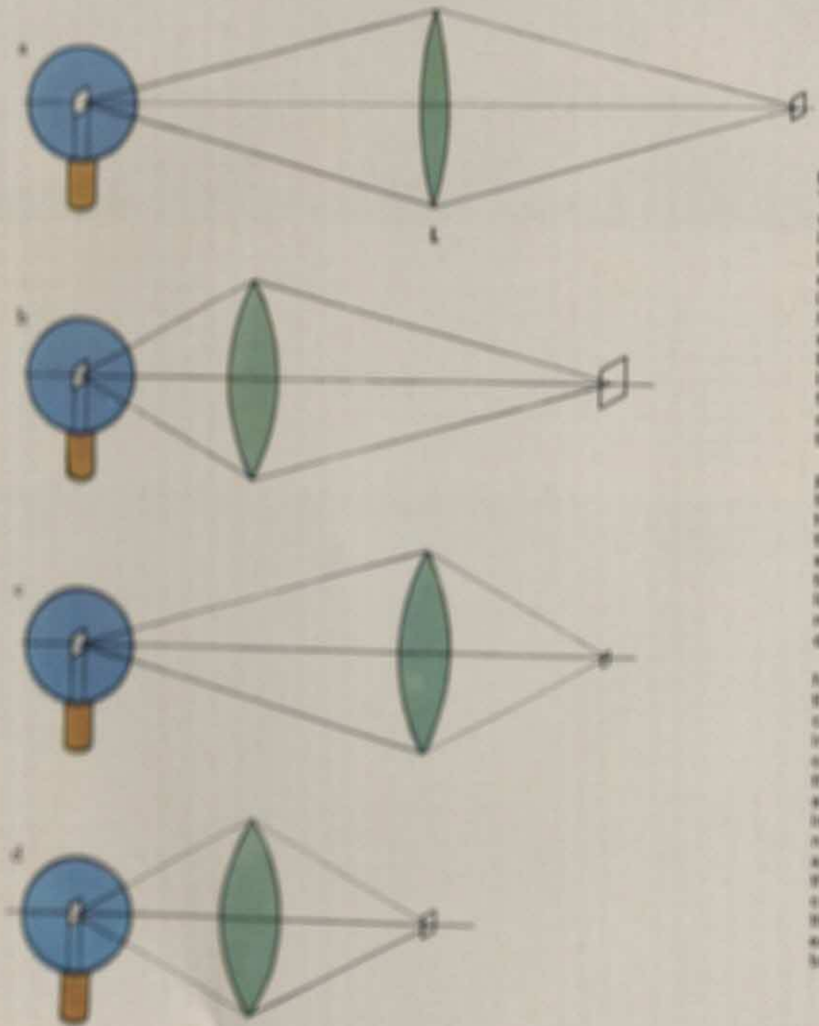
SOURCES OF LIGHT | important physical characteristics

Light and the phenomena connected with it have long been the subjects of investigation, but important work on the theory of light actually began with the electromagnetic wave theory of the Scottish mathematician and physicist James Maxwell (1873) and Einstein's special theory of relativity (1905). The most recent and satisfactory theory of light is included in modern quantum physics, first formulated by Einstein.

Popularly, light is often described in terms of its effects on the human eye. However, certain light rays (infrared and ultraviolet) cannot be seen, but are used in medical and other applications. Chemical processes may be greatly influenced by light since photons in the visible region (0.0004 to 0.0008 [illegible] have energies of about the same magnitude as those involved in the exchange [illegible] arrangement of atoms among molecules. Light is also used as a concentrated source of energy and as a carrier in communications.

A special branch of physics called photometry deals with the measurement of the relative intensity of light sources. In

1

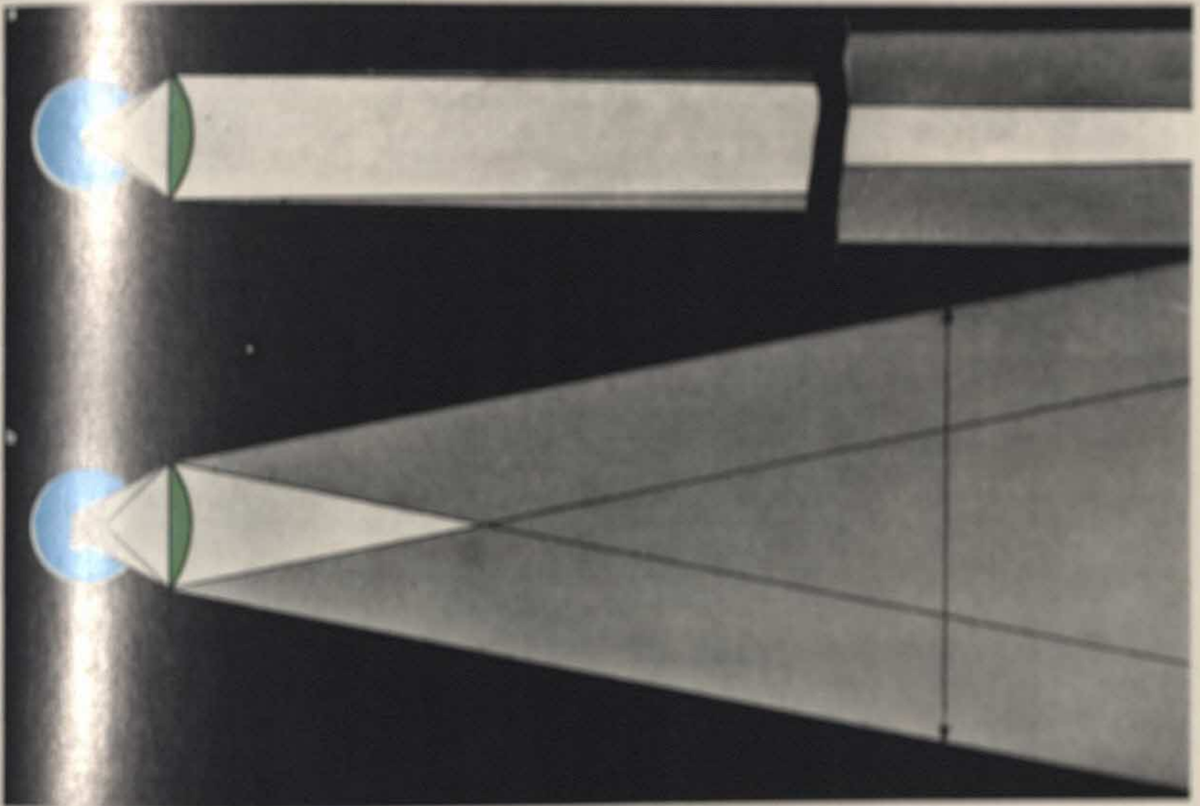


BRIGHTNESS OF OBJECTS AND IMAGES—The optical system shown consists of a tungsten filament lamp bulb that radiates equally in all directions. In illustration 1a, converging lens L, when struck by a portion of the light emitted by the filament, produces an image of it. Since the distance between filament and lens and image are equal, the image has the same dimensions as the filament. However, the brightness of the image is less than that of the light source because the lens accepts only a fraction of the radiation dispersed. The amount of light emitted that will be concentrated onto the image depends on the lens.

In illustration 1b, a high-quality lens is placed close to the source in such a position that the distance between lens and image is twice that between lens and filament. Thus, the lens can accept light radiations from a solid angle four times as great as in illustration 1a. In effect, this increase in accepted light doubles the image size, but the brightness remains the same as before since the dispersion angle is also increased.

In illustration 1c, the lens is moved away from the source so that the distance between the lens and the image is shortened. In this case, the lens concentrates light onto a smaller image but accepts it from a smaller portion of the angle. Thus, the image is no brighter than before. The solution is to place the lens as close as possible to both the source and image, as shown in illustration 1d. The brightness is then four times that of the other examples, but no greater than that emitted by the source. It is theoretically possible to increase image brightness to an extent greater than the source, but this requires complex and expensive equipment, and the increase in brightness that results is minimal.

2



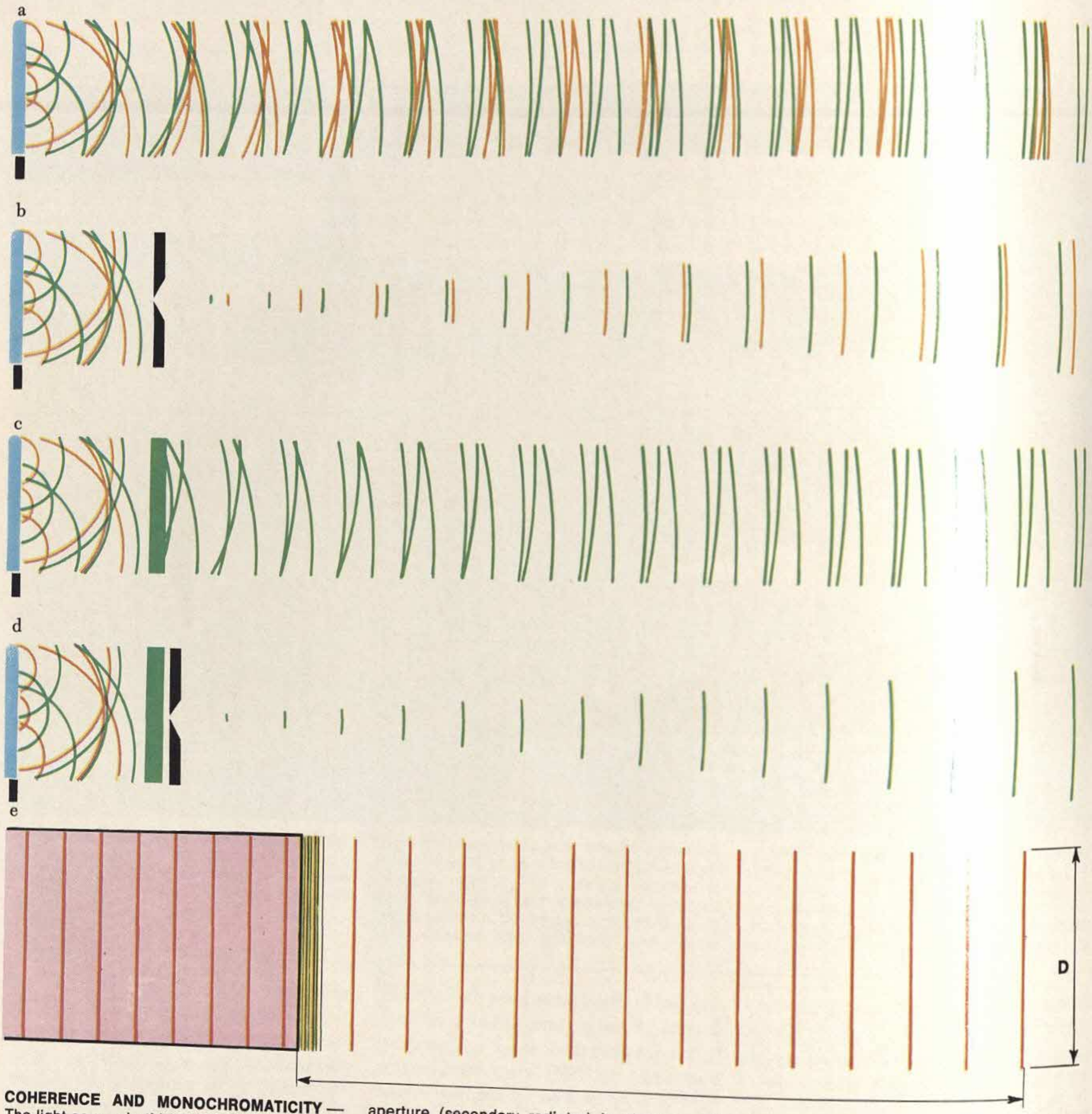
PARALLEL BEAMS AND LARGE SOURCES — The light source in this illustration determines how parallel the beam will be at infinity. In both cases, the beams are not perfectly parallel but converge to some point on the horizon. This is much more evident in illustration 2b, where the lens is quite large and the beam more widely divergent than in illustration 2a. In addition, any lens will produce some diffraction in the beam and in the resulting image. These limitations are inherent in all beams and explain why even the most powerful light sources (lighthouses, for example) always use a lens with a long focal length and a light source as concentrated and bright as possible, such as a voltaic arc.

addition to bodies that emit light and those that reflect it, other types of bodies are best described by the way in which they affect light (air, glass, tissue, lead).

Light sources are indispensable to many physical experiments. The refraction of light by a lens to form an image is involved in optical instruments such as the camera, microscope, telescope, and the human eye. When the "color of light" is mentioned, it means the radiations emitted by the source have similar wavelengths. Because transparent substances (glass, for example) do not uniformly refract light of the various wavelengths, light may be broken up into a rainbow-colored band of its component colors—violet, blue, green, yellow, orange, and red—by passing it through a prism.

Light beams can be produced with superintensities (lasers, for light amplification by the stimulated emission of radiation). The laser produces an extremely intense, pencil-sized, sharply directed beam of light. It is invaluable in the study of atomic phenomena, and in the study of the properties of electromagnetic radiation. The immense electric fields produced when the beam of coherent light is focused to a fine point cause an interaction of the light with the molecular vibrations in materials, producing shifts of the light frequency.

3



COHERENCE AND MONOCHROMATICITY— The light source in this illustration is large and the radiations are characterized by two spectral lines. Such radiations might be produced by a gas-filled tube made fluorescent by an electrical discharge. In Illustrations 3a through 3d, the light waves emitted by such a source are changed as the result of different conditions. First, a diaphragm with a small aperture is placed in front of the light source. As a result the wave fronts are spread out from the aperture (secondary radiator) in circular and concentric fashion and are easily distinguished according to wavelength. When the diaphragm is replaced by a green filter, only the green light waves are emitted (monochromatic radiation); but these waves are incoherent and the wave fronts vibrate out of phase with each other. By combining filter and diaphragm, the light waves are both filtered and cohered. The wave fronts are thus all in phase and equally separated since only one of the two monochromatic wavelengths has been selected.

In Illustration 3e, the light source (laser) emits radiations that are parallel and equidistant in a cone that will increase by only a minute amount as the distance increases. For example, at a distance ten thousand times the beam diameter, the beam does not even double. Thus, the laser produces light beams more parallel than those from ordinary light sources even when these sources are filtered or cohered.

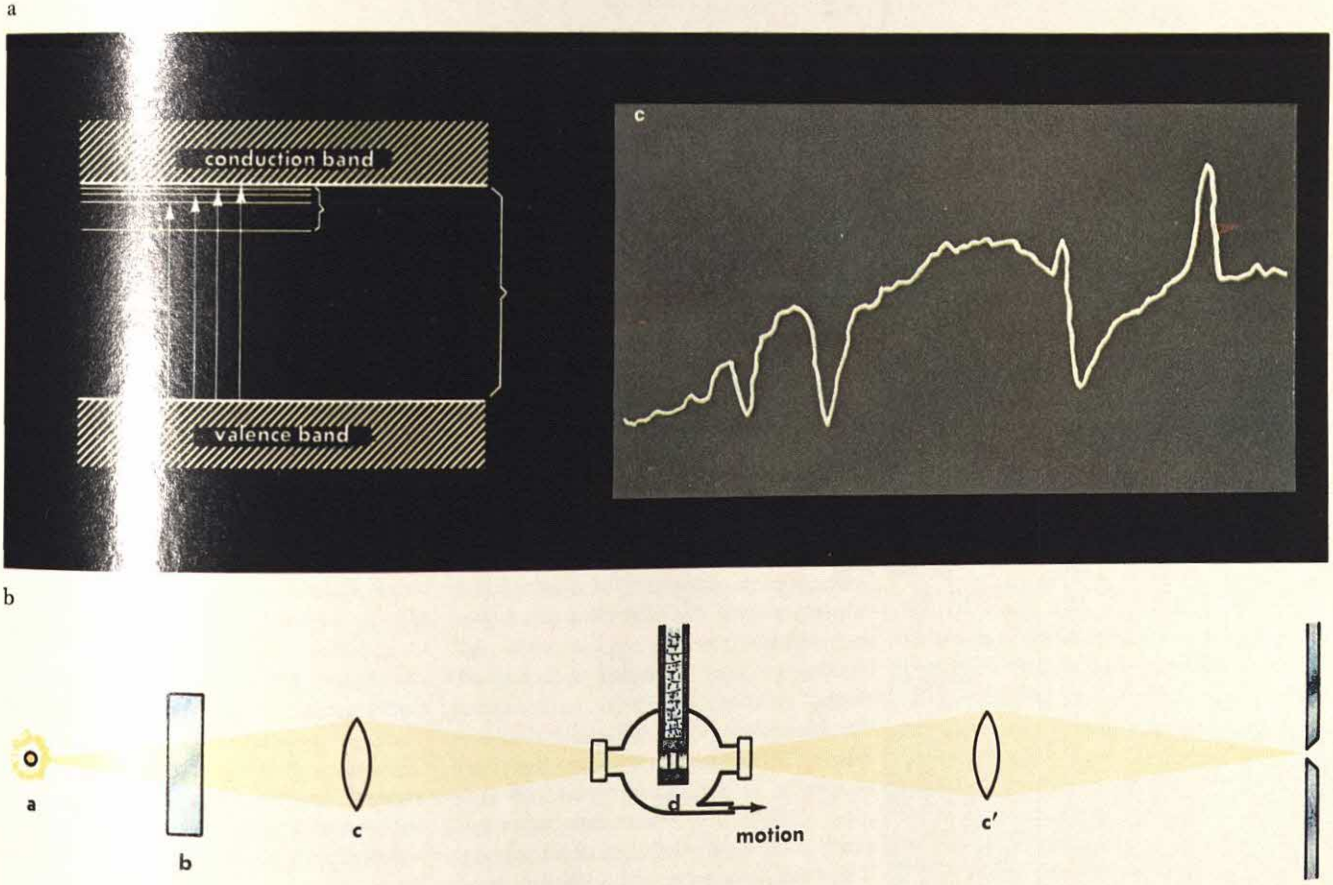
EXCITONS AND LUMINESCENCES

phosphorescent substances

The term *luminescence* refers to a discharge of energy in the form of light radiated from a substance charged by an external energy source. Prefixes are added to the term to indicate the type of energy charging the substance that emits the light. Thus electroluminescence is luminescence caused by electrical energy; and photoluminescence is caused by light energy. Other types of energy, depending on the circumstances, can also serve to charge solids.

Luminescence, a phenomenon that first aroused the curiosity of researchers well over a century ago, remains a subject of research in physics laboratories. Luminescence phenomena can be categorized either on the basis of the sources of energy that produce them, or on the basis of the duration of the time interval between charging and emission of light.

1



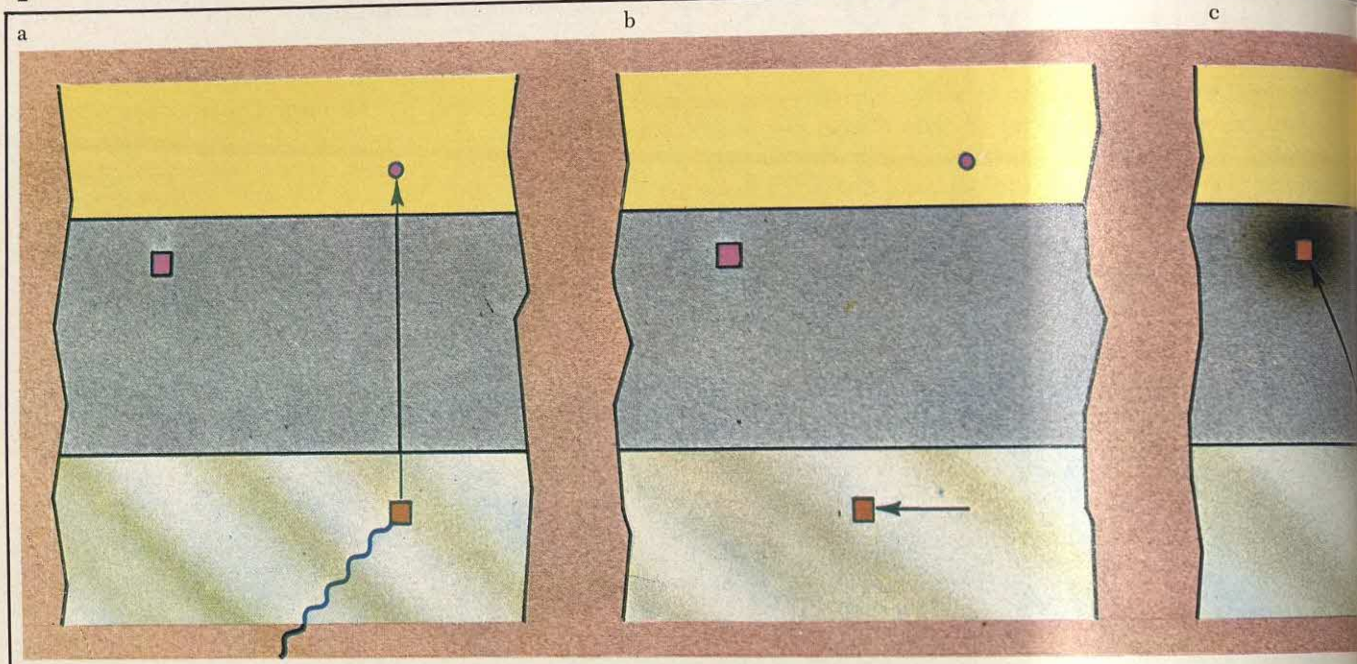
THE EXCITON—A sufficiently energized photon penetrating certain crystals has the capacity to produce an electron-hole pair. The crystal in which the electron-hole pair is so produced is neutral. The electron, however, can combine with the hole because of the mutual attraction of their opposite charges. In a certain sense, the electron-hole pair behaves like the nucleus and the electrons of a hydrogen atom. In the case of the hydrogen atom, the electron can occupy different, well-differentiated energy states and can, therefore, emit a spectrum. However, an electron-hole pair can only absorb energy. Moreover, the energy is absorbed during the formation of the electron-hole pair.

Illustration 1a shows the energy levels of the electron-hole pair. The electron is normally at the level of the valence energy. The arrival of a photon can raise its energy to a higher energy level. If the energy of the photon is great enough, the electron may even occupy the conduction band.

The experiment shown in Illustration 1b demonstrates that the energy states of the electron with respect to the hole are similar to the energy states of the electron and proton of the hydrogen atom. A source of light **a**, such as a high-pressure mercury vapor lamp, emits radiations that are filtered by means of a small basin **b** filled with water to eliminate a large part of the infrared radiation that would heat the specimen. A condenser **c** concentrates the radiation on the specimen **d.** The specimen is kept inside a cryostat because heat affects the outcome of the experiment. Finally, a second condenser **c′** concentrates the radiation in the entrance slit of a spectrophotometer.

Illustration 1c is the spectrophotometric record of the radiation. It includes five energy levels of the hydrogen type.

2



THE MODEL OF A PHOSPHORESCENT SUBSTANCE — The phenomenon of phosphorescence is not easy to understand because each phosphorescent substance is excited and discharged differently. The following example is rather typical and can serve as a guide to an intuitive understanding of how other substances phosphoresce. The substance in this example is a zinc sulfide activated with a monovalent impurity.

A photon (indicated by a wavy line in Illustration 2a) enters a crystal and produces a hole (shown by the red square) and a free electron. This action of the photon is called excitation. The electron migrates into the conduction band.

The hole also migrates within the crystal (Illustration 2b). However, it is less mobile than

The temporal means of characterizing luminescent phenomena is based on the time needed for an electron to carry out a transition within a free atom. To be more specific, an atom of hydrogen that has been excited by a collision is charged with energy. After a very brief time, the atom will be discharged by the emission of a photon. The time interval involved is on the order of 10^{-8} seconds. Luminescence phenomena whose radiation time is of this order of magnitude are known as fluorescence. A typical example of fluorescence is the light emitted by the crystalline dye fluorescein when it is excited by ultraviolet radiations. Phenomena in which radiation is emitted after a longer time—from millionths of a second to several minutes—are known as phosphorescence. Cadmium sulfide activated by silver is an example of a typically phosphorescent substance.

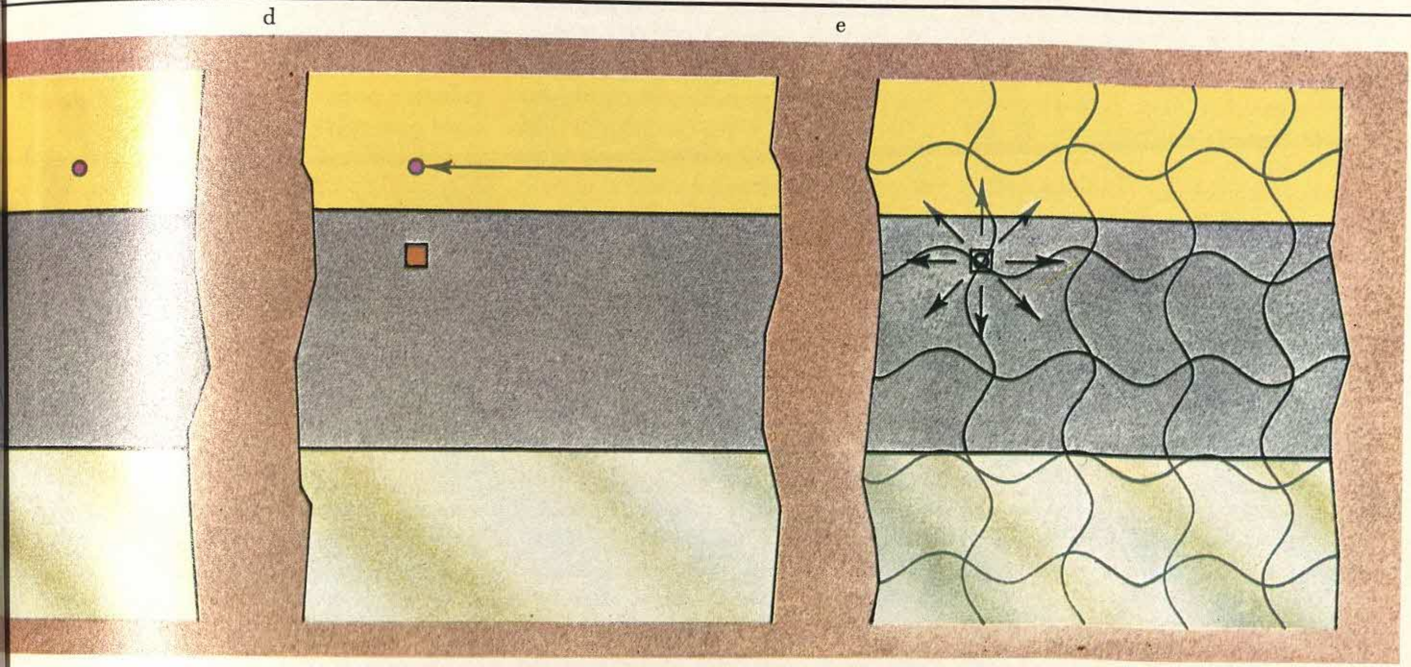
PRESENT APPLICATIONS

Luminescent substances have been used for several decades in many branches of science and technology and in the manufacture of domestic appliances.

One of the first uses of luminescent substances was the detection of x-rays and cathode rays by oscilloscopes. An oscilloscope tube is coated with a luminescent coating known as a phosphor. The phosphor, which is usually zinc sulfide with about one part per ten thousand of copper, or a mixture of zinc and cadmium sulfide, may also contain other impurities that can delay its fluorescence. The amount and type of impurities used in the phosphor depend on the purpose for which the oscilloscope is to be used. If phenomena of short duration are to be observed, the phosphor should leave only a short luminescent trace on the oscilloscope screen. For phenomena of longer duration, a longer trace is better. The length of time during which the trace of the phosphor remains luminous and clearly visible after excitation has stopped is called the persistence of the phosphor.

An even more common use of luminescent substances is the fluorescent lamp. The mercury vapor with which these lamps are filled emits an intense spectral line at 2,537 Å (Angstroms) when excited. The radiation, however, is not visible, and a fluorescent substance is used to transform the radiation into visible light. The complex halophosphate of calcium that is used fluoresces with different colors, depending on the substance by which it is activated. Antimony produces a predominantly blue light; manganese, an orange light. These two elements, when properly balanced, produce white light.

Fluorescent substances are also used in the picture tubes of television receivers, which are similar to oscilloscopes in some respects. Picture tubes require phosphors of very low persistence because the image must change very frequently. Every time a new image appears on the screen, the previous image must be canceled or decayed completely. Television picture tubes must also have a perfectly balanced white light. In view of the very brief period of decay necessary for television reception, the problem of what combination of phosphors to use is very complex.



the electron and moves more slowly.

In the course of its motion, the hole approaches an impurity in the crystal and is attracted to it (Illustration 2c). The hole gains an electron from the impurity. This impurity is called the activator because when it combines with the hole, light is emitted. However, the hole combined with the impurity does not constitute a neutral whole; it will still attract a free electron. The free electron formed from the photon continues to move through the crystal (Illustration 2d).

Finally, the free electron is captured by the impurity from the hole and relieved of an electron (Illustration 2e). Energy is again released during this process but not in the form of light—in the form of infrared radiations and in the form of vibrations of the crystal lattice.

Picture tubes for color television represent an even more complex problem in this respect. They require the use of as many as three different types of phosphors that must be so chosen that the mixture of their emissions will give white light. Any imbalance in the combination of the phosphors used could cause an undesirable alteration in the color reception of the set. The production and use of phosphors, as seen in the applications discussed here, is a very complex and highly refined process.

FUTURE APPLICATIONS

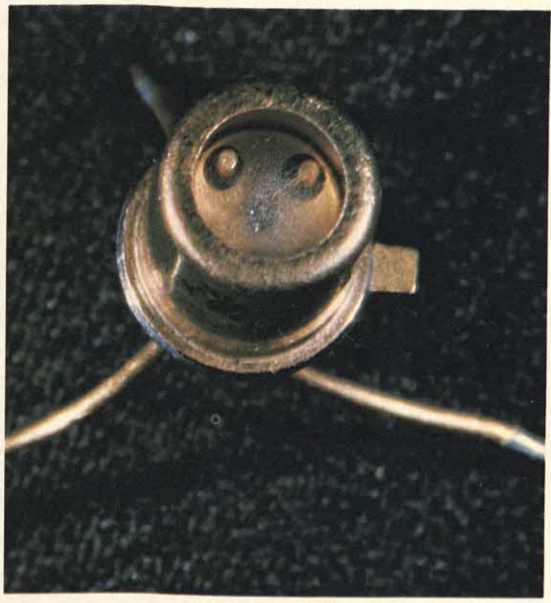
Possible future applications of luminescence have been foreshadowed by recent advances. For example, some compounds become emitters of light when crossed by an ionizing particle, as in scintillation counters. These compounds (usually halogenides of alkali metals activated with thallium or solid solutions of anthracene, naphthalene, and terphenyl) are used today for the detection of ionizing particles in experiments in nuclear physics.

In luminescent processes, two ratios are important. The first is the conversion ratio (the ratio of the induced energy to the light energy); the second is the ratio of the light energy to the induced energy. The latter is a measure of the efficiency of the phosphor. All phosphors used today are relatively inefficient. The efficiency of electroluminescent substances is even lower than that of phosphors. Electroluminescent substances in use today are semiconductors that have been treated so that the electric energy induced in them is converted directly into light. However, it is hoped that studies now in progress will eventually make possible lamps in which the emitting element is a semiconductor connected directly to the supply network.

Picture tubes in television receivers may someday be replaced by flat electroluminescent screens of any desired size. However, the only advanced application of electroluminescent substances is the construction of photoemitters, small lamps that emit light when a voltage is applied to them.

THE PHOTOEMITTER—The photoemitter, when subjected to the proper voltage, emits a beam of radiation with an aperture of 15°. The wavelength of these radiations is about 1.8 μm (microns) and corresponds to the wavelength to which photodiodes and phototransistors are sensitive.

3



LIGHT AND CRYSTALS | instruments for analyzing polarized light

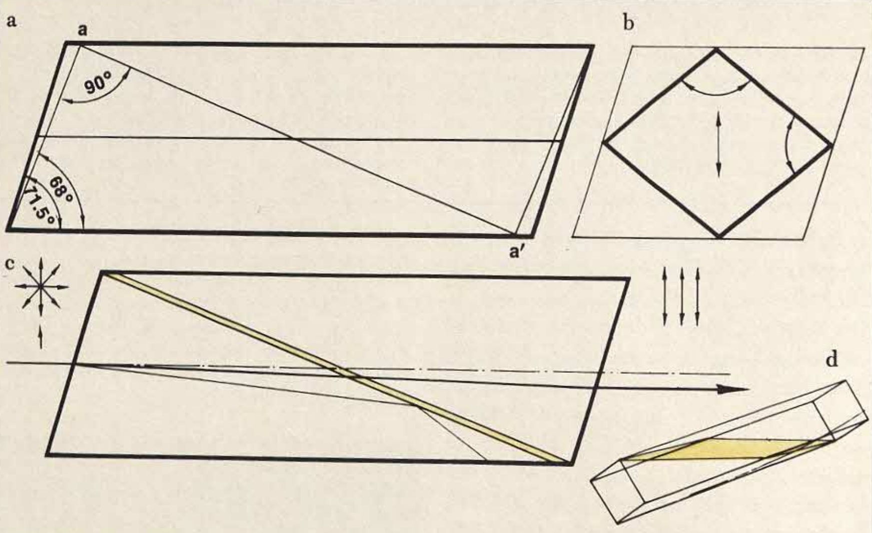
To throw light on a concept means, in common parlance, to clarify it. Light itself seems so simple. However, light is composed of various rays that act and react in complex ways. A case in point is what happens when light is "thrown" on crystals.

In the phenomenon of double refraction, a transparent substance will display a different index of refraction, depending on the direction of the light's propagation. This phenomenon has many applications because the two rays into which a beam of normal light divides when it enters a crystal are polarized at right angles.

The results obtained by the experiments that follow can also be obtained by means of other arrangements that make use of the double refraction properties of crystals; the crystal systems, however, display properties that are greater than those of synthetic substances. To judge a polarizing system, the two polarizers are arranged so that the second extinguishes the light polarized by the first. When crossed by a beam of light, the first polarizer should allow 50 percent of the normal light to pass, selecting only the polarized component. The second polarizer, on the other hand, should completely stop the beam of light that has crossed the first.

This extinction will occur if both polarizers are perfect, which means that the light that strikes the second polarizer is stopped completely. However, this is perfection. Actually, it will be found that, owing to very small imperfections of the crystal or some minor defects (a scratch, for example) on the surfaces crossed by the light, the pair of polarizers does not stop the whole of the beam; rather, the polarizers allow the passage of a few parts per million of the light that has struck the train of prisms (the pair of polarizer prisms arranged in line on the same optical path). The polarizing power obtained with artificial systems made of plastic materials is much less. This is only one of the applications of birefringent crystals; others exhibit properties such as dichroism, in which the two rays entering certain crystals are absorbed differently. Another most interesting property is accidental double refraction: the birefringence displayed by substances such as glass or celluloid which are isotropic under normal conditions but become birefringent when subjected to mechanical stresses. These substances are used to analyze the stresses in models of machinery and structures.

1



NATURAL POLARIZATION—A prism that polarizes light can be produced by exploiting the double refraction properties of many substances. Calcite is usually chosen because it is easy to work with; also, it can still be found in large and perfectly transparent pieces. A ray of light entering one face of a Nicol prism at a certain angle divides into two rays that follow different paths, provided the prism has been cut from the crystal at a suitable angle as shown here. The prism is cut along a diagonal plane and the two parts are glued together. The entering beam becomes decomposed into two rays. One is totally reflected by this glued plane in such a way that it is absorbed by the walls of the prism. The second ray, on the other hand, continues on its way without disturbance and becomes polarized.

The heavy lines in Illustration 1a represent the outline of a natural rhombohedron of calcite. The angle at the base is 71.5°. From this a second rhombohedron is cut with more oblique faces; the angle is 68°. The outline of this new rhombohedron is shown with a lighter line. It is then cut along the plane **aa'**, which forms a right angle with the faces.

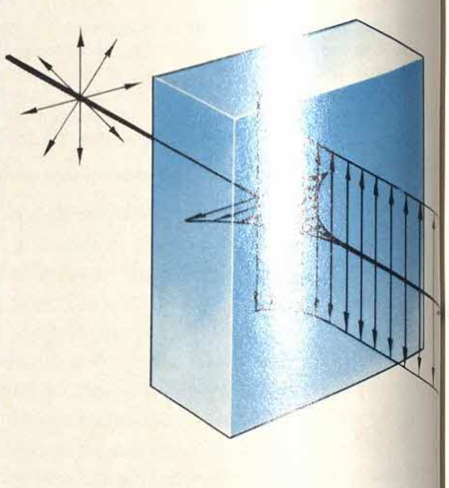
At this point the sides of the crystal are also cut until the profile of the rhombohedron is reduced as shown. In Illustration 1b the section of the entrance face is shown as seen from the direction from which the light is coming.

Illustration 1c shows how the glued surface of the prism is arranged; the glued joint is made of Canada balsam (a purified coniferous resin). The glue has an index of refraction between the ordinary and the extraordinary ray of the calcite. Therefore, if a ray strikes the calcite-balsam plane at a suitable angle (as shown), the ordinary ray will be totally reflected.

The outer surface of the prism is blackened to enable it to absorb the ray it is desired to extinguish.

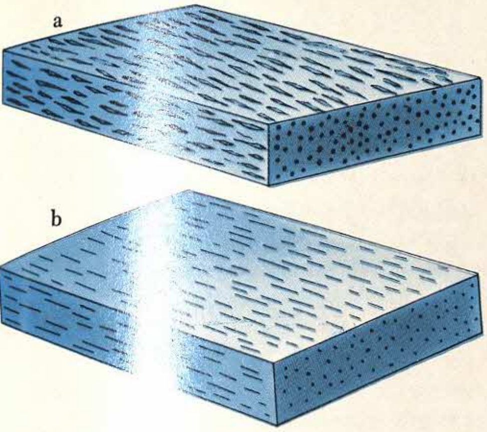
Illustration 1d shows how a Nicol prism is formed. Long and relatively thin, it is capable of completely polarizing only a very thin beam of light. This liability is common to all the polarizing systems, even the most perfect dichroic ones.

2



THE PHENOMENON OF DICHROISM—This illustration shows, greatly enlarged, a lamina of a crystal displaying the property of dichroism. A normal ray of light entering the crystal is divided into two components; one is polarized in one plane and the other in a perpendicular plane. This occurs in all birefringent crystals; in most cases, however, both rays will proceed on their way through the crystal while undergoing a small absorption, which is equal for both the rays. (Even the most limpid and transparent crystals have a small absorption.)

Other types of crystal, on the other hand, have the characteristic of being strongly absorbent for one of the rays but only slightly absorbent for the other. If the lamina of this crystal is thick enough, one of the two polarized beams will be practically extinguished and will not pass out of the crystal. The other beam will pass almost undisturbed. A crystal with this property is said to be dichroic. Tourmaline, for example, because it possesses the property of dichroism, can be used to produce polarizers of light because a small thickness is quite sufficient to absorb one of the two rays of light.



3

SYNTHETIC DICHROIC POLARIZERS — A method for producing synthetic polarizers having properties similar to those of natural prisms was devised in 1934 by the American inventor Edwin Land. These polarizers are based on the principle of dichroism. First it is necessary to produce a powder of herapathite crystals. These crystals are dichroic, but they display this phenomenon only when they are observed in the right direction. When immersed in a solution of plastic material, they remain in suspension. This suspension is then made to flow into a channel of suitable form that causes the crystals to become oriented; subsequently the solvent is evaporated and a plate is obtained. This plate displays the property of polarizing just as though it were a slice of dichroic crystal cut in the right direction (Illustration 3a).

A more economical method of producing a synthetic polarizing sheet, polyvinyl alcohol, was discovered after the Second World War. A sheet of this substance (Illustration 3b) becomes birefringent when it is subjected to a permanent elongation by means of a tensile stress. However, it is not dichroic: the two beams cross the sheet without suffering any absorption. If the substance is colored with iodine, the sheet becomes dichroic, but it also absorbs part of the polarized light, reducing the transparency of the polarizer. In many instances, however, it is convenient to use the polyvinyl alcohol polarizer, especially since its cost is very low compared to the cost of natural crystals.

4

RETARDING LAMINAE — The phenomenon of double refraction offers the possibility of carrying out some special experiments in regard to polarized light. In fact, a birefringent lamina can be used to rotate the plane of polarization of a beam of light through 45° or 90° or through any other angle. Consider a block of a birefringent substance oriented as shown in Illustration 4a. A crystalline lamina of parallelepiped form is cut from it. The light striking this lamina along the direction of the arrow **L** is propagated in the direction **O.**

Illustration 4b is the lamina as observed from **O,** with the arrows indicating the direction of the light. Since the lamina is birefringent, there are two types of wave fronts. The first is made up of spheres, the second of ellipsoids (seen here in section). Consequently, the extraordinary ray is propagated more rapidly than the ordinary ray. The dotted line indicates the direction of the optical axis.

On the other hand, Illustration 4c shows the ordinary ray preceding the extraordinary one. The first is drawn with the plane of polarization vertical; the second, with the plane of polarization horizontal.

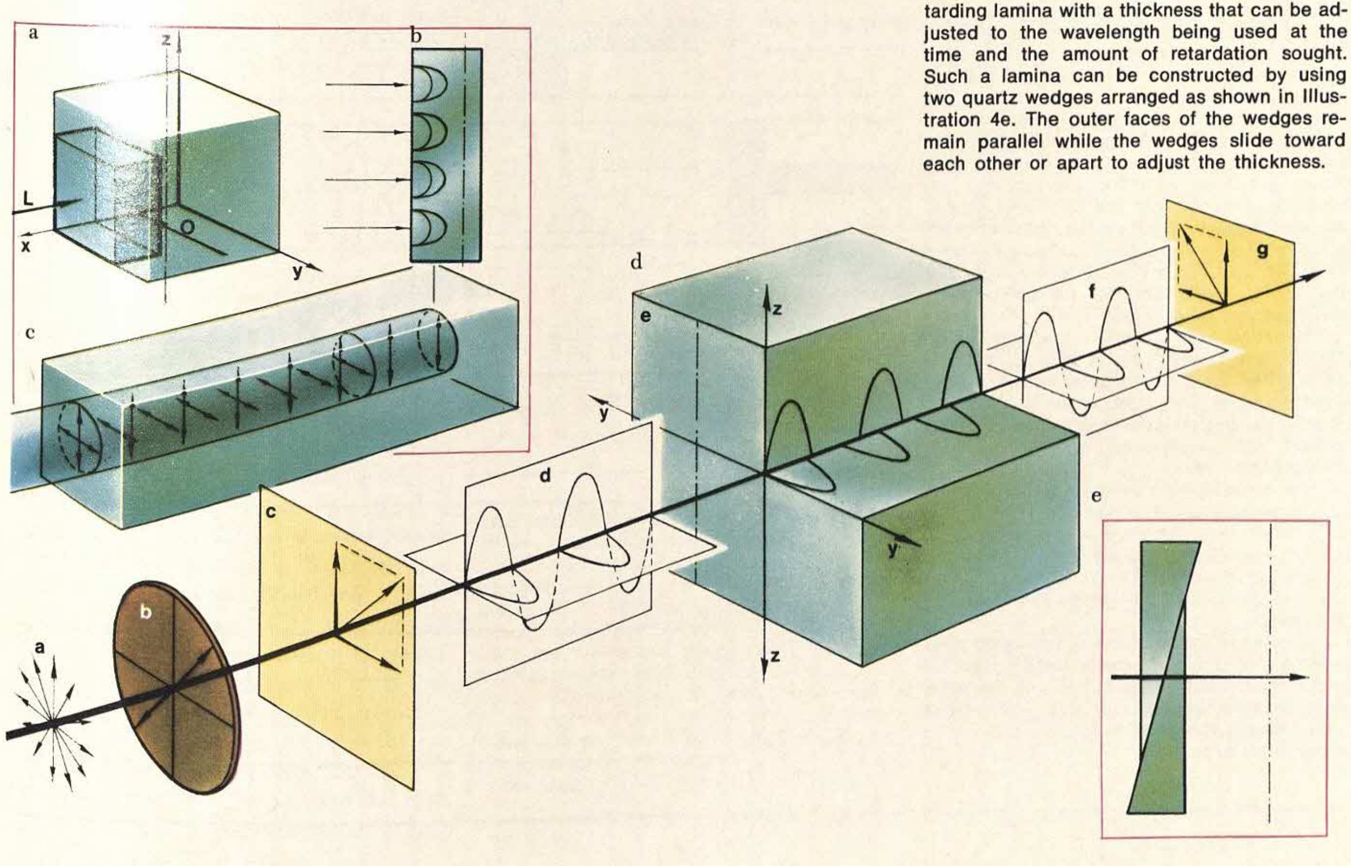
The object of Illustration 4d is to show how a lamina of a birefringent substance, cut as in Illustration 4a, acts on a ray of light. A normal beam of light **a** falls on a polarizer **b,** whose direction of polarization is inclined 45° to the right with respect to the vertical.

An indication at **c** shows that the polarization vector can be decomposed into two directions, into vertical and horizontal vectors, so that the electromagnetic wave can be represented as the result of two component waves **d.** The optical axis of the lamina **e** of a birefringent substance is indicated by the dotted line.

The vertically polarized light is propagated through this lamina at a greater speed than the horizontally polarized light. The waves in the horizontal plane are thus retarded with respect to those in the vertical plane.

In **f** it is shown that the light that has emerged from the lamina (which has retarded the horizontal component by exactly half a wave with respect to the vertical one) is polarized in a different manner from the light that entered the lamina. In fact, **g,** the sum of the two components, gives an electric vector inclining to the left rather than to the right. The direction of the polarization has actually been changed by 90°. A birefringent lamina that has the characteristic of retarding one of the polarized beams by half a wave length while passing through it is called a half-wave lamina. Its thickness obviously depends on the wavelength being considered, because a birefringent substance, like any other transparent substance, has an index of refraction that depends on the wavelength of the light passing through it. For example, a half-wave lamina made of mica will have a thickness of about 72 μm for the red radiation of hydrogen. It is relatively easy to construct a lamina of the required thickness from mica, which splits into thin sheets. A lamina of some other material would have to be ground down to a thickness previously calculated on the basis of refraction of the substance and amount of retardation required.

Certain optical experiments call for a retarding lamina with a thickness that can be adjusted to the wavelength being used at the time and the amount of retardation sought. Such a lamina can be constructed by using two quartz wedges arranged as shown in Illustration 4e. The outer faces of the wedges remain parallel while the wedges slide toward each other or apart to adjust the thickness.



THE INTENSITY OF SOUND

how sound is generated and how its intensity is measured

The rustling of the leaves of a tree moved by a light breeze and the roar of a jet engine are opposite extremes of the sound intensity the human ear can perceive. The ear cannot hear sounds weaker than the light rustling of leaves, while noises louder than that of a jet engine are not heard as sounds but felt as painful sensations. Between these two extremes is a range of perceptible sounds whose intensity can be analyzed.

The ear distinguishes several characteristics of such sounds, including the direction from which the sound is coming, its pitch, and (within certain limits) what kind of source has emitted the sound.

To understand what intensity means, as applied to sound, the nature of sound and how sounds are generated must be considered. If a piston is moved back and forth rapidly inside a cylinder, it alternately compresses and rarefies the air in the cylinder. The points at which the air is compressed or rarefied do not remain fixed, but move toward the open end of the cylinder. If the ear is placed against this end of the cylinder, a sound is heard. The longer and more rapid the stroke of the piston, the greater will be the compression and subsequent rarefaction of the air, and the louder (or more intense) the sound.

The expression "a sound strong enough to split the eardrums" is not an exaggeration. If the pressure of a sound wave is great enough, the membrane that closes the outer eardrum will be torn by the thrust of the compressed air in the sound wave. This process is analogous to that of a gas pushing a piston inside a cylinder.

Sounds of ordinary intensity (for example, a conversation or the noise of traffic in the street) set up sound waves of very low pressure. The pressure of sound waves set up by leaves rustling in the wind exceeds atmospheric pressure by only 0.0000000002 percent. Yet this seemingly insignificant additional pressure is quite sufficient to be perceived by the human ear.

A person standing some 50 m (about 160 ft) behind a jet engine may hear a sound whose pressure is about 3,000 percent higher than atmospheric pressure.

Since sound waves contain energy, a body that absorbs them will become heated. This heating is relatively small,

1

HOW A SOUND IS GENERATED—A tube filled with a gas is closed at one end by a piston (Illustration 1a). The piston can slide freely inside the tube, which is filled with air. The air is represented by means of the molecules of the gases that compose it (the molecules are actually much smaller and closer together than they appear in the illustration). It is assumed that the density of the molecules is approximately the same throughout the tube.

In Illustration 1b, the piston is moved rapidly to the right. The density of molecules near the piston is increased, thus increasing the pressure in this area. This compressed air tends to expand into the air next to it. Thus, in a short interval of time, the high density zone of air molecules moves toward the right.

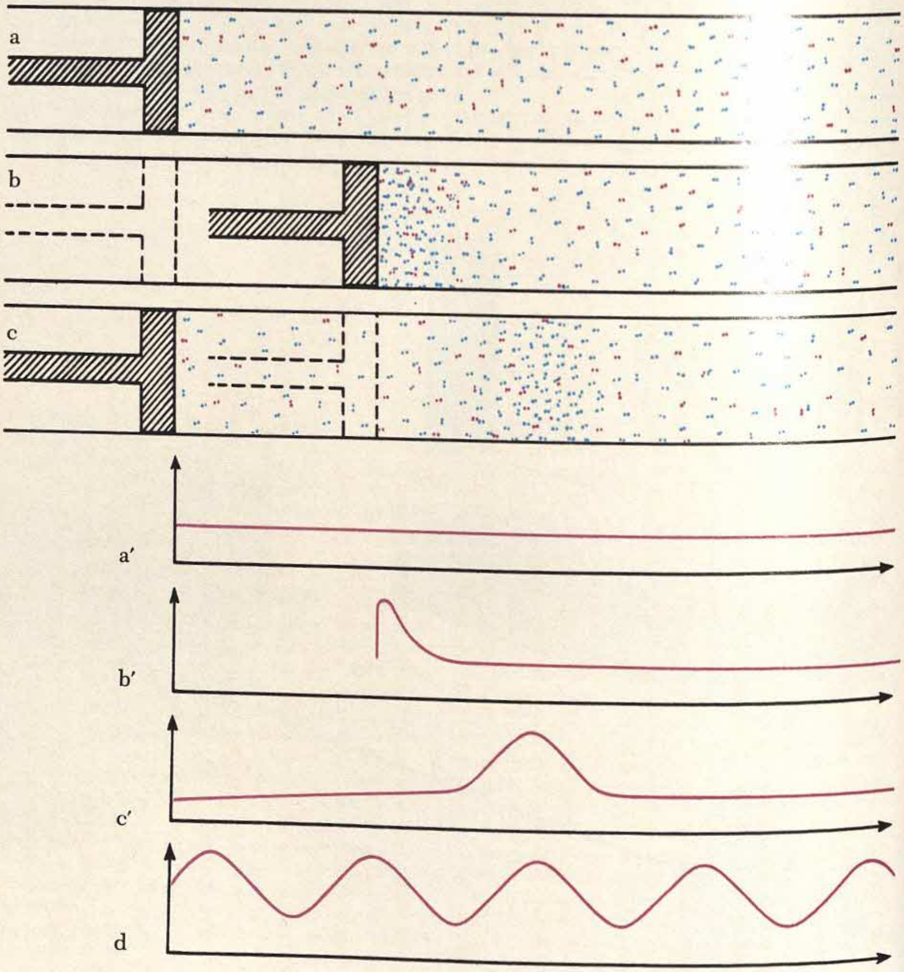
In Illustration 1c, the piston is withdrawn to its initial position, and some of the air molecules rush into the empty space the withdrawn piston has left in the tube. However, the air molecules moving to the right have caused the compression of the air they have encountered on their way, and this compressed zone of air, in turn, causes compression even further to the right. Therefore the compression has "become propagated."

Illustration 1a′ represents the pressure of the air (ordinate values) at points along the tube. Since the pressure is assumed uniform (and equal to atmospheric pressure) at all points along the tube before the piston is moved, the pressure is represented by a straight line.

After the piston is moved rapidly to the right, the pressure at points immediately to the right is greater than atmospheric pressure (Illustration 1b′). Further to the right, however, it remains unchanged. This is represented by a peak in the position where the pressure has increased.

After the piston is returned to its initial position, some molecules have rushed into the empty space left by the piston, but the pressure remains lower in this area. The area of compression has, meanwhile, moved to the right (Illustration 1c′).

Continuous moving of the piston back and forth causes successive compressions and rarefactions of the air inside the tube. The compressions and rarefactions are spoken of as causing a "pressure wave" and are represented by a wavy line (Illustration 1d).



but its quantitative measurement gives a measure of the energy contained by the sound waves. A square centimeter of a surface placed close to the exhaust of a jet engine will have to remain there for twelve hours before the quantity of energy it receives in the form of sound waves will be sufficient to raise the temperature of a cubic centimeter of water by one degree. (The energy from the sound alone is considered; any heating due to exhaust or other elements is neglected.)

MEASURING SOUND INTENSITY: THE DECIBEL

The intensity of a sound is the amount of energy it will transmit to a surface area of one square centimeter in one second. This standard of measurement is used in physics experiments. In practice, a simpler system of measurement is used. Since it is usually meaningless to measure the intensities of sounds that cannot be perceived by the human ear, the 0 point of the scale is chosen to represent the intensity of sound that is just at the threshold of audibility. A sound level 10 times more intense (that is, a sound that sends 10 times more energy to a given surface than the barely audible sound) is indicated by the number 10 on this scale. A sound that is 100 times more intensive than the barely audible sound is given a value of 20. A sound 1,000 times as intense is given a value of 30. This scale is thus logarithmic. The units are called decibels (or tenths of bels, after the American physicist Alexander Graham Bell). Thus the intensity of a sound in decibels is given by the following equation:

$$db = 10 \log (I/I_0),$$

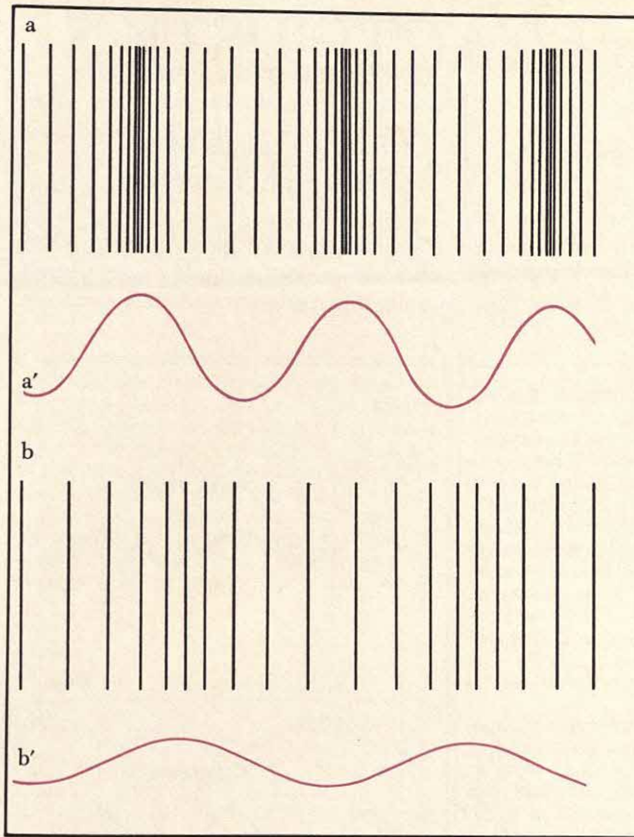
where db is the value in decibels, (I/I_0) is the ratio of the intensity of the sound being measured to the barely audible sound that is used as a base (and given a value of 0 db as stated above), and log presents the logarithm to the base 10.

Sounds sufficiently intense to cause pain to the human ear have an intensity of at least 120 decibels. This means that their intensity is about a trillion times greater than that of barely audible sounds.

Water transmits sounds like air, but the speed of transmission is much greater —1,441 m/sec (about 4,750 ft/sec) in water compared to 332 m/sec (about 1,100 ft/sec) in air. The reception of sound and the measurement of sound intensities underwater are carried out with special instruments. For example, sonar (an acronym derived from the expression "*so*und *na*vigation *r*anging") is an instrument that uses acoustic means to localize submerged objects and to survey the ocean floor. Sonar consists essentially of two elements: a transmitter that is placed in the keel of the ship and that sends sound pulses toward the sea floor, and a receiver that registers these pulses if they are reflected by some obstacle.

The sonar instrument measures the time that elapses between the transmission of the pulse and the reception of its echo. The speed of sound in water is known; therefore, the depth of the obstacle can be calculated.

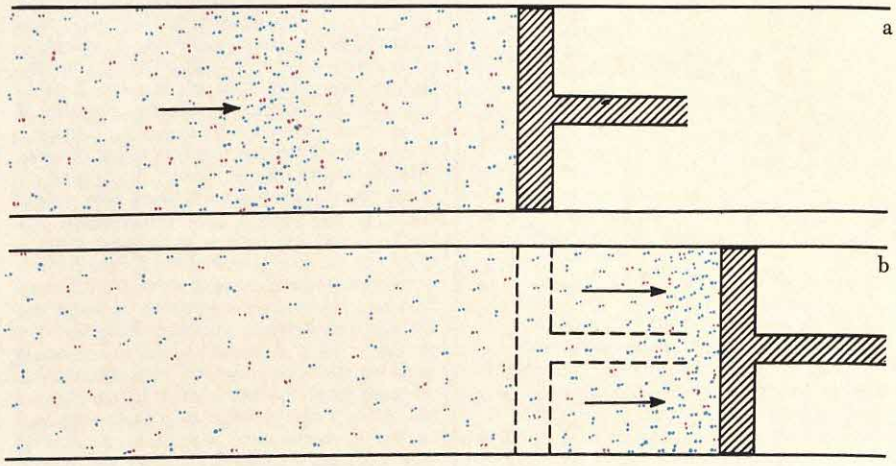
2



INTENSE AND WEAK SOUNDS—Illustration 2a represents an intense sound. The vertical lines represent the density of the air. They are closely spaced where the air is compressed and far apart where it is rarefied. This is represented by a curve (Illustration 2a′) that rises and falls sharply.

Illustration 2b represents a weak sound that corresponds to relatively little compression and rarefaction of the air. The variations of pressure are shown in Illustration 2b′, a curve that rises and falls less sharply.

3



THE ENERGY OF SOUND WAVES—Illustration 3a shows a cylinder filled with air in which a sound wave is being propagated. At the point where the air is most dense, the molecules are moving to the right. However, these molecules do not travel far because as they move to the right, they collide with nearby molecules, push them to the right, and themselves bounce backward. The molecules they have struck move to the right in turn, collide with other molecules, and so forth. Thus a sound wave contains energy for two reasons: because the air in the wave is compressed (and can perform work as it expands); and because it contains molecules in motion.

How the energy of a sound wave is observed. When the compressed air shown in Illustration 3a reaches the piston that is free to slide inside the cylinder, it pushes the piston to the right (Illustration 3b). The piston is displaced a certain distance to the right due to the force exerted on it. The energy needed for this work, which is equal simply to the force exerted times the distance moved, came from the sound wave.

THE CHARACTERISTICS OF SOUND

different aspects of the same phenomenon

Why can the sound of the human voice be caressing, while the sound of a nearby jet aircraft is unpleasant to the ear? Why is the former called *sound* and the latter *noise?* Why does a musical note played on one instrument sound different from the same note played on another instrument even when the intensity of the two notes is the same? These and other dif-

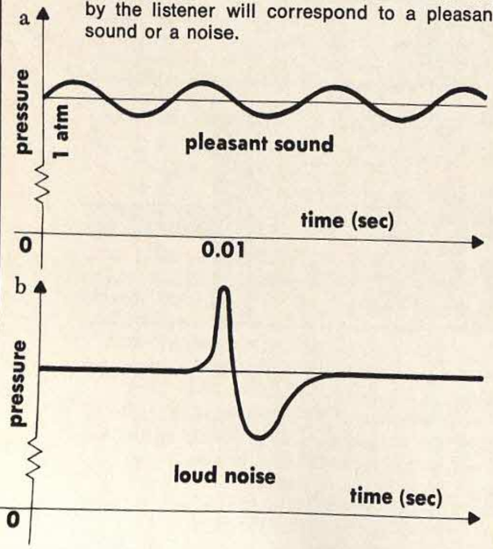
1

THE PRESSURE DIAGRAM OF A SOUND WAVE—Changes in atmospheric pressure are often the harbingers of changes in the weather. These changes are measured by a barometer (from the Greek word *barus,* meaning "heavy"), an instrument that registers variations of atmospheric pressure with considerable accuracy.

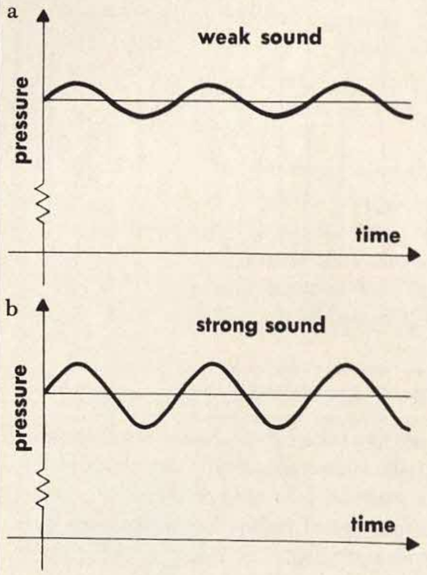
The pressure of the air changes when sound waves travel through it, but a barometer does not succeed in registering these variations; this is due not only to the fact that the pressure variations are usually quite small, but also because they occur with such rapidity that the instrument is unable to respond quickly enough.

Nevertheless, certain instruments are capable of measuring the pressure variations that result from the passage of sound: some of these transform the pressure variations into weak electric current that is multiplied by an amplifier. When a musical note is played on an instrument, any instrument, the ear hears the note because it responds to a rapid alternation of compressions and decompressions of the air. This alternation can be well represented by a sinusoid (Illustration 1a). The line representing the axis of the sinusoid is placed at a height that corresponds to the value of the average atmospheric pressure. The graph is a pressure diagram of the sound wave. A pressure variation such as this produces an impression of sound that can be quite pleasing.

A pressure diagram that is characterized by a sudden increase of pressure followed by a slow return to normal atmospheric pressure corresponds to a sensation of noise (Illustration 1b). Only by observing the pressure diagram of a sound source is it possible to predict approximately whether the sensation felt by the listener will correspond to a pleasant sound or a noise.



2



THE INTENSITY OF SOUND—Intensity is an extremely important and distinctive characteristic of sound. The intensity of a sound is understood to mean what is commonly described by the word loudness: a sound is easily heard when it is loud or strong, but difficult to hear when it is soft or weak. A pressure diagram indicates what the intensity of a sound consists of. For example, Illustration 2a is the pressure diagram of a sound of weak intensity emitted by a musical instrument; Illustration 2b represents the same note played loudly. In the second case the pressure diagram still has the form of a sinusoid, but the amplitude is now greater than it was before. Intense sounds are those in which the pressure rises and falls a long way above or below the average atmospheric pressure. However, the phrase "a long way" still refers to extremely small variations of pressure; if the atmospheric pressure is of the order of 1 kg/cm² (about 14.2 lb/in.²), the increase of pressure (as well as the corresponding decrease) associated with the most intense sound that the human ear is capable of hearing will amount to a little less than 0.3 g/cm² (about 0.137 lb/in.²). This is the pressure increase in the immedate neighborhood of a jet engine operated at maximum thrust (Illustration 2c).

In a quiet environment when human ears can hear the noise made by a mosquito (Illustration 2d), the pressure variations of this level of sound intensity do not exceed more than two ten-billionths of a kilogram per square centimeter, or two ten-billionths of the atmospheric pressure. The molecules of air that transmit this extremely weak sound move forward and backward in the direction of the propagation of the sound, just as they do in the case of all sound waves: again, the sound alternately compresses and rarefies the air. In the case of a sound as weak as this, the movement (forward and backward) of the molecules amounts to one billionth of a centimeter. In other words, the alternating displacement of the molecule amounts to no more than a tenth of its diameter. The average wavelength of visible light amounts to as little as half a thousandth of a millimeter: the sound wave made by a flying mosquito causes an alternating displacement of air molecules that is fifty thousand times smaller than this wavelength.

c

d

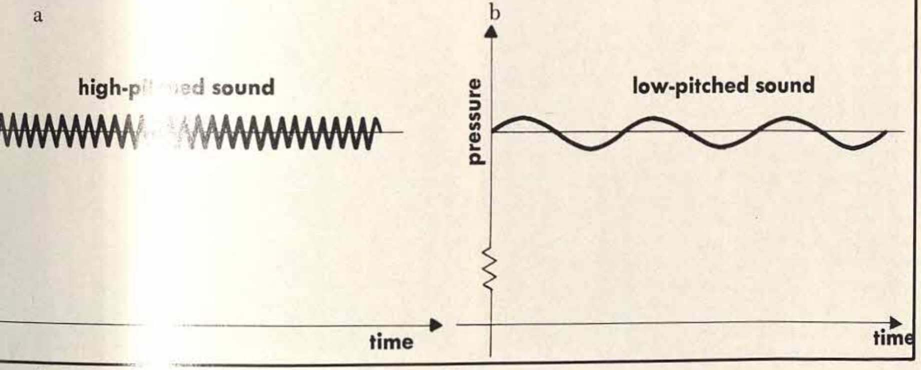
ferences between various types of sounds are determined by some of the characteristics of sound waves. An understanding of the basic nature of sound answers the questions formulated.

The sense of hearing is extremely important to man: it takes second place only to sight. The spoken word together with the sense of hearing is the prime medium for communication among human beings; music is a source of pleasant sensations that the ear transmits to the brain. However, the need for transmitting the human voice over long distances and for reproducing and transmitting music raises special acoustical problems, some of which are discussed in this article.

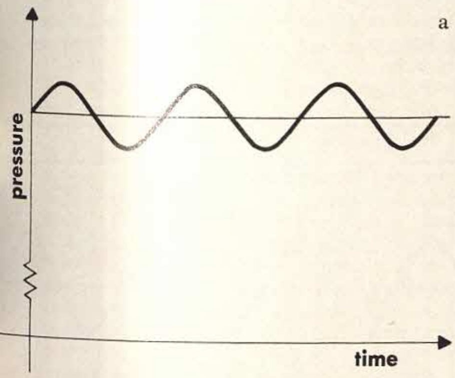
3

THE PITCH OF SOUND—The pitch of a sound is determined by its wavelength. A sound is high in pitch if its pressure diagram consists of a sinusoid with many closely spaced peaks, that is, a sinusoid having a brief period (Illustration 3a). Low-pitched sounds, on the other hand, have a pressure diagram consisting of a long, flat sinusoid (Illustration 3b). The pressure diagram represents the pressure variations of the air with respect to time and at a particular point: the pressure variations then become propagated from that point, spreading into space at an established and definite speed. The speed of propagation in air at 0° C (32° F) and at standard atmospheric pressure is about 331 m/sec (about 1,087 ft/sec).

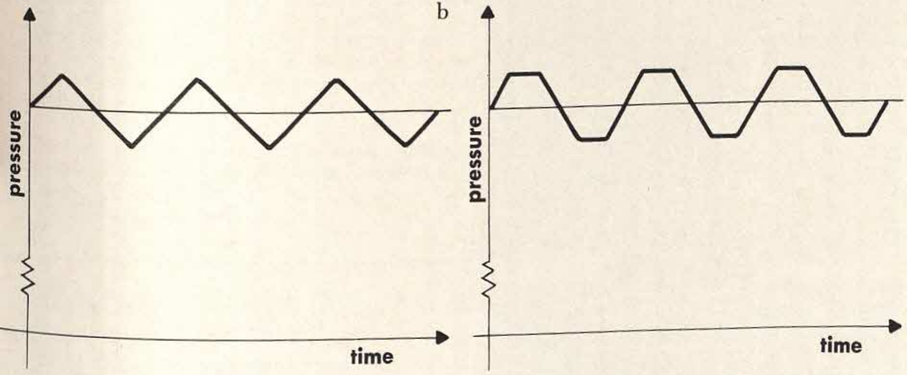
The exact numerical relation between the wavelength (or sound frequency) and the pitch of sound form the principal subject matter of musical acoustics (see p. 76 of this volume).



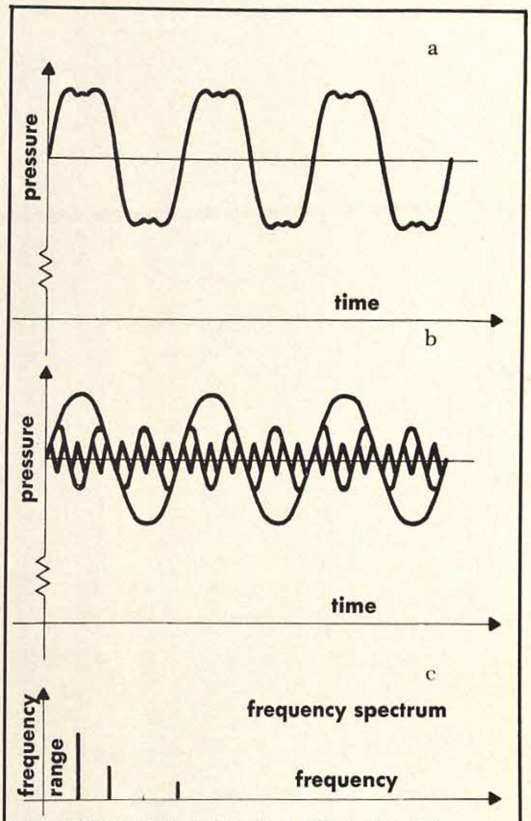
4



THE TIMBRE OF SOUND—A note sung by the human voice does not sound like the same note played on a musical instrument, even if the notes have the same pitch and intensity; and the note played on one instrument does not sound like the same note struck on a different musical instrument. The difference is *timbre,* a characteristic produced by the form of the wave in the pressure diagram. The pressure diagrams refer to sounds identical only in pitch and intensity: Illustration 4a corresponds to a sound with a perfectly sinusoidal pressure diagram, reproducible only with a special instrument such as a tuning fork. Illustrations 4b and 4c represent the sounds of two different instruments.



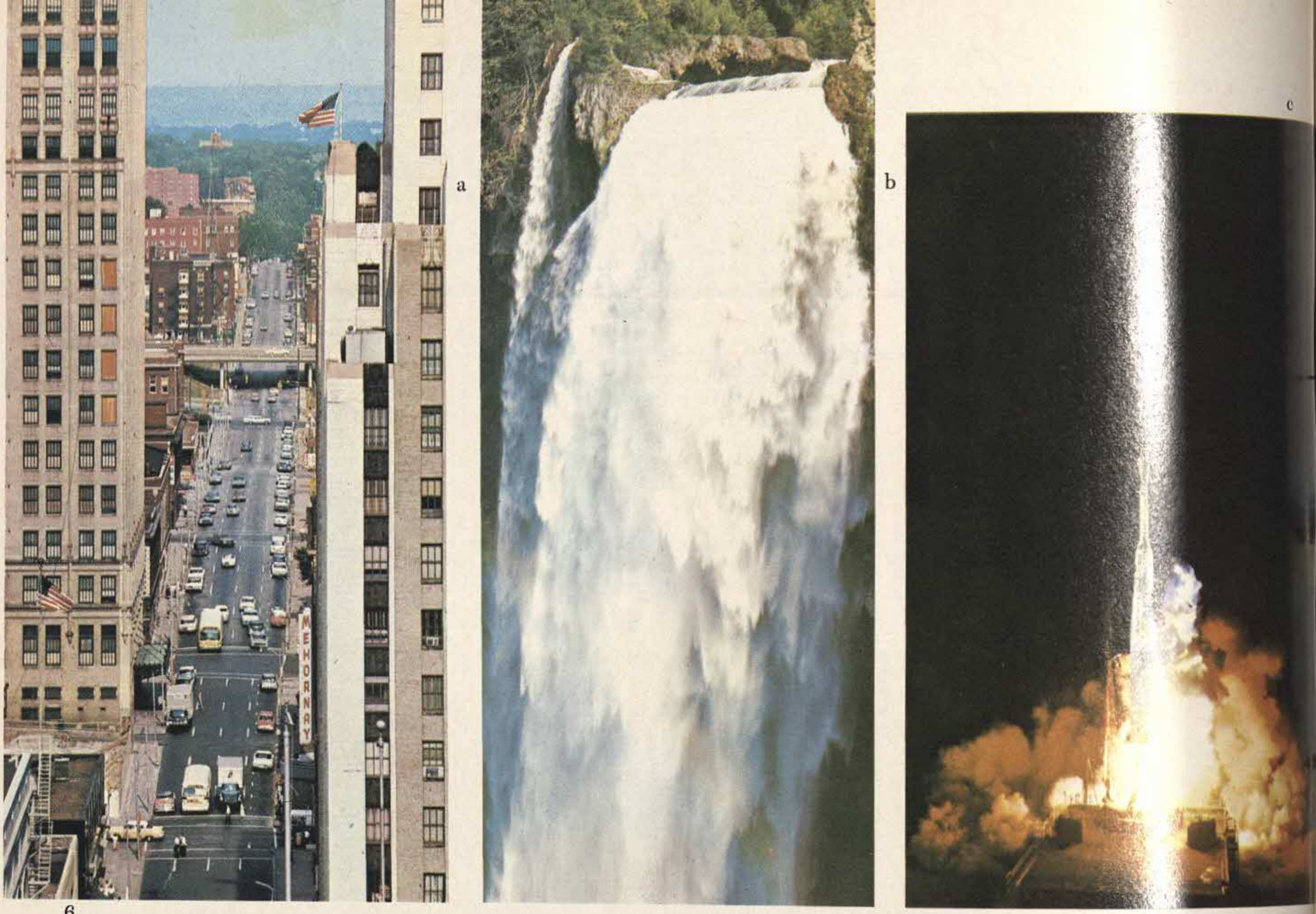
5



THE FREQUENCY SPECTRUM OF A SOUND AND ITS TIMBRE—Periodic phenomena such as those represented by pressure diagrams of sound can be analyzed by means of a method discovered by the French mathematician Jean Baptiste Fourier. The almost sinusoidal form of the pressure diagram can be imagined as the average form of sinusoids of different frequencies, each of which has a distinctive amplitude of its own. For example, in the case of the sound emitted by a violin, the pressure diagram is only approximately sinusoidal; it is really considerably deformed. This form can be obtained by summing the three sinusoids shown in Illustration 5b. The superposition of these three distinct sounds, which the source emits simultaneously and in phase with each other, gives rise to the wave form shown in Illustration 5a. In phase means that the various points of the different sinusoids are always summed in the same manner at regular intervals of time; if they were emitted by three different sources, they would not give the sensation of a single sound and would be perceived as three sounds with distinct timbres.

The frequency spectrum of a sound is the diagram that shows the relative abundance of the different frequencies present in a compound sound. In the case of the sound emitted by a violin, it consists of three vertical segments, each of which represents the intensity of one of the three sounds that make up the compound sound. The term *spectrum* is derived from the analogy with the optical spectrum and the mass spectrum.

The human ear is extremely sensitive to timbre: the famous violins constructed by Stradivarius and the Amati families at Cremona produced the same notes as all other violins, but their timbre was such as to confer on any kind of music a quality that was quite unmistakable and that remains unforgettable for those who have heard it.



6

WHITE NOISE—The rumble of traffic as heard by someone high above it, the roar of a waterfall, the thunder of exhaust gases rushing violently out of a rocket nozzle: all these (Illustrations 6a, 6b, and 6c) have a wholly distinctive timbre. In fact, all these noises consist of equal quantities of all the possible and imaginable frequencies rather than of a few frequencies as in the case of the violin. The curve representing the distribution of these many various frequencies is, therefore, a horizontal straight line: by analogy with optics, where a curve of this type corresponds to the spectrum of white light, this sound spectrum is called the spectrum of white noise or noise that contains all the frequencies and all the possible pitches of sound.

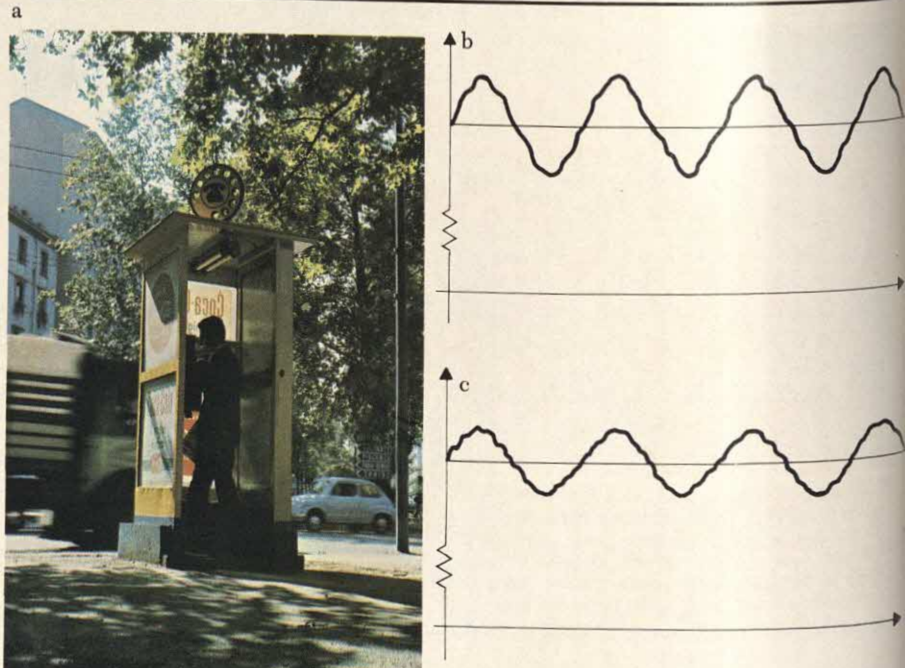
7

THE PURITY OF SOUND—In a telephone booth (Illustration 7a) on a busy city street, listening to the voice of a person speaking at the other end of the line, the voice may sound indistinct. This is partly due to the fact that the white noise of traffic reaches the ears at the same time as the voice is heard. What other transformations may this voice have undergone in the process of being transmitted? The appropriate pressure diagrams answer the question.

Illustration 7b shows the form of the wave of the voice at a short distance from the mouth of the person who is speaking (assuming that he is speaking in a silent environment).

This pressure diagram of the sound wave will already be transformed by the telephone circuit; when the voice arrives at the ear of the listener, its intensity has been reduced and the high and low frequencies of its Fourier sum have been cut off.

Illustration 7c is the pressure diagram of the voice heard at the other end of the line; the curve includes the white noise. All sounds heard are mixed with white noise; the more a sound is characterized by the absence of white noise, the purer it is said to be (because it is less altered). The rustling sound produced by a phonograph pickup when playing a record where the grooves have been worn by prolonged use is another form of white noise.



THE DOPPLER EFFECT | movement alters frequency

Its horn blaring, a diesel locomotive speeds toward a person standing by the track. To that person's ears, the sound rises in pitch as the train approaches, reaching a crescendo toward the point where he stands. As soon as the locomotive passes, the sound of the horn drops in pitch. The frequency of the sound waves is different because of the relative motion of their source—the horn on the locomotive—toward and from the fixed position of the observer. With proper instruments, the same effect can be observed with other types of emanations, such as visible light, ultraviolet waves, radio waves, and infrared radiation. This phenomenon is of tremendous value because of its scientific and practical applications. Many radiation sources are in motion relative to an observer: on an atomic scale, on the human scale, and on an astronomical scale.

The change in frequency of radiations emitted because of the relative motion of the radiating source and the position of an observer is called the Doppler effect, after the Austrian mathematician C. J. Doppler. It is a phenomenon easily understood, with many significant examples in everyday life.

1

A HYDRAULIC ANALOGY—Many problems in physics concerned with waves can be explained most simply by a hydraulic analogy. This illustration shows a section of beach with ocean waves breaking on it. An observer can count the number of waves coming in each minute—for example, 10 waves a minute, or a wave every 6 seconds. If the same observer puts out in a fast boat, perpendicular to the incoming surf, and counts the waves again, he will count many more than 10 waves a minute. If he turns the boat about and heads back in toward the beach—counting the waves again—he will find fewer waves per minute than he counted from a fixed position on the beach. If he slows the boat to the exact speed of the incoming waves, his count will be zero; he and the boat will be either in a trough or on a crest, passing no waves. Seamen have practiced this technique for centuries when met with violent storms, pointing the bows of their ships in the same direction the waves are moving. This direction frequently coincides with the direction of the wind; fewer waves are encountered and their force is less violent against the ship.

The scientific point of this example is that motion relative to the source of wave propagation causes a change in the frequency of the waves encountered by the observer. This change is the Doppler effect.

2

THE DOPPLER EFFECT—The principles of the Doppler effect are shown in this illustration. A complete explanation of this effect would require a discussion of relativity; a practical understanding, however, can be obtained from the examples and comparisons described below. It should also be remembered that the Doppler effect occurs with both sound and electromagnetic waves, but there is an additional complication with sound waves. The speed of sound becomes added to (or subtracted from) the speed of the air within which the sound is propagated. The speed of light, of course, is quite independent of the reference system.

To begin, examine the wave presented in Illustration 2a. **A** is the source; **B** is the position of an observer. They are separated by a certain distance. **A** radiates at a frequency whereby **B** receives the first wave just as **A** is emitting the fifth wave. This fifth wave reaches **B** a certain time after the observer has received the first wave.

In Illustration 2b, the situation is changed. **A** is emitting at the same frequency as before, but is moving toward **B.** The five waves emitted are compressed into the shorter distance **A'B.** Since **B** receives those waves at shorter intervals than before, they occur at a greater frequency to him and appear to have a shorter wavelength.

In Illustration 2c, **B** is moving toward **A.** He will receive the last wave sooner than he did in the first example and the frequency (waves per minute) will appear to have increased.

Illustration 2d shows both **A** and **B** moving toward each other. The distance between them is decreased from **AB** to **A'B'.** The frequency of emissions observed by **B** will thus be even more rapid.

A more complicated case is shown in Illustration 2e. Here, **A** and **B** are close together, but **B** receives the radiation by means of a reflection from mirror **S.** If neither **A** nor **B** change position, **A** will emit 10 waves before **B** receives the first one.

Expanding on Illustration 2e, assume that the mirror **S** is in movement toward **A** and **B;** that it goes from **S** to **S'** during the time **A** is emitting 10 waves. The waves will arrive at **B** more frequently than before and thus will appear shorter.

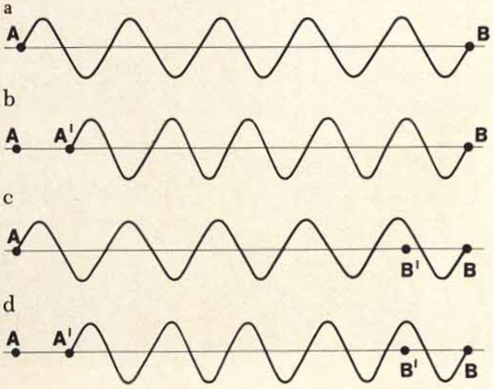
Illustration 2f shows a different situation. **A** (the radiation source) moves from **A** to **A'** while it is emitting. The frequency of waves received by **B** will therefore be increased.

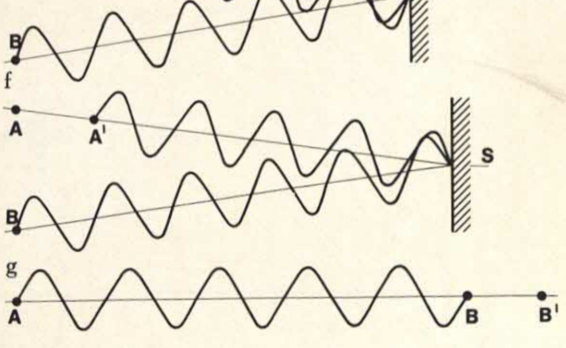
In Illustration 2g, **B** moves away from **A** (from **B** to **B'**) during wave propagation. If an observer receives the first wave at **B** and the last wave at **B',** the frequency of the waves will appear to be reduced, and their wavelength to be longer.

For electromagnetic waves, the formula must take into account the theory of relativity: the speed of light is constant. Let the frequency observed be f_o, the frequency emitted by the source be f_e, let v be the relative speed between the source and the observer, and c be the speed of light. The formula given is $f_o = f_e \sqrt{(c + v)/(c - v)}$. Moreover, if v is much smaller than c, the above formula is approximated by $f_o = f_e\ (1 + v/c)$. The latter formula is usually used, since the speeds considered are always much smaller than the speed of light.

The speed of sound, however, is not independent of the motion of the medium in which it is propagated and transmitted. If V is the speed of sound in the air, W the speed of the wind in the direction of propagation, v_o the speed of the observer, and v_e the speed of the source, the formula for the observed frequency is

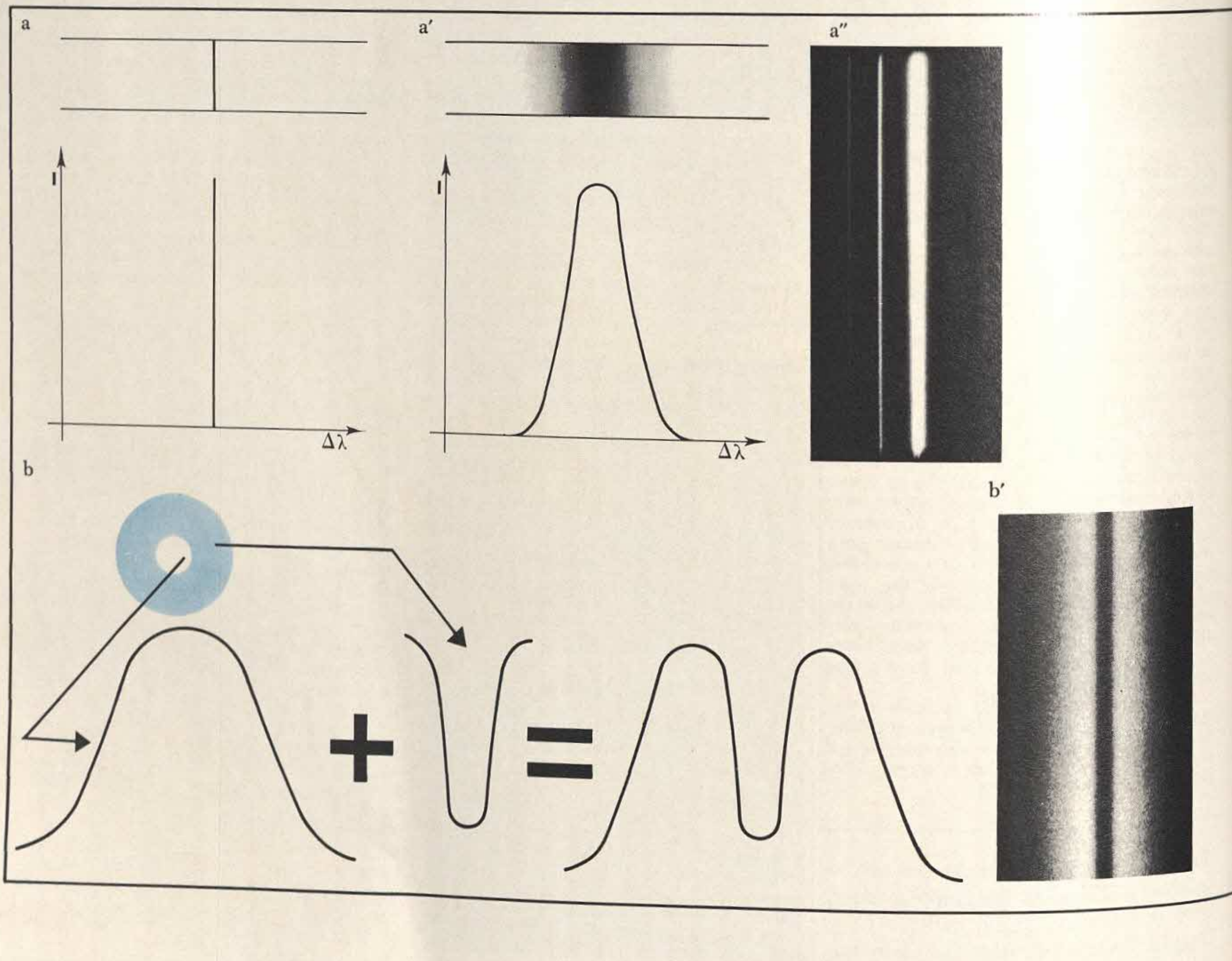
$$f_o = f_e\ /(V + W - v_o)/(V + W - v_e).$$





3

4

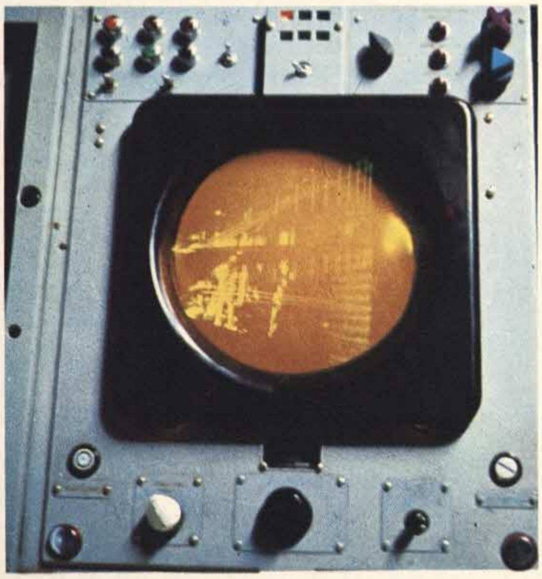


ACOUSTIC APPLICATIONS—A common example of the Doppler effect in acoustics is the variation in the pitch of a sound as the relative distance between the source and the observer varies. Such an example was cited earlier. Consider now the situation shown in Illustration 3a. The horn of a car normally emits two well-defined notes. The pitch of these notes is higher when the vehicle is approaching an observer and lower when it is moving away. The greatest variation is at the moment it passes the observer. Illustrations 2b and 2g illustrate graphically why this occurs.

The Doppler effect is used in certain kinds of ultrasonic detectors. Shown in Illustration 3b is a sonar (sound *na*vigation *r*anging) set used for locating underwater objects. The sonar emits an ultrasonic beam directed toward the ocean floor, and receives the echo. An obstacle between the instrument and the ocean floor reflects an echo earlier than the one received from the ocean bottom. If that object moves, the ultrasonic beam is reflected at a different frequency—higher if the object is approaching and lower if the object is receding.

THE DOPPLER EFFECT AND MICROWAVES—A radar (*ra*dio *d*etecting *a*nd *r*anging) set emits microwaves and receives echoes wherever these waves strike an object. A beam of electrons is then directed toward the screen of a cathode ray tube, at a point corresponding to the position of the object that caused the echo. In a busy airport, incoming and outgoing aircraft are controlled by means of radar, but there are many more obstacles to the radar beams than aircraft alone: echoes are picked up from surrounding buildings, pylons, surrounding high ground, and other stationary objects. To identify moving objects, modern radars make an analysis of the frequency of microwave echoes and emit electrons to the cathode ray tube only if the echoes are received from moving objects. Radar sets are also designed for tracking of satellites, signaling only those echoes reflecting from an object with a speed corresponding to a selected satellite.

5



SPECTROSCOPIC APPLICATIONS — One of the most important applications of the Doppler effect is measurement of the radial velocity of stars and nebulas. The speed of these celestial objects ranges from several kilometers per second to several thousand kilometers per second. A similar application is that of studying the spectral emission lines of objects in space.

Theoretically, the emission of light by atoms on a single wavelength should appear as a single and well-defined line when analyzed by a spectroscope. If the source of this light and the observer were at rest relative to one another, the wavelength should be clearly defined. If this were true in application, the spectral line would appear as shown in the upper part of Illustration 4a—as a clear, thin line. The intensity distribution would accordingly be observed as shown in the lower part of Illustration 4a—as a clear, thin line.

In reality, all atoms are in some state of thermal agitation. Some are moving toward the observer; some are moving away from him. As a result, what is actually seen by means of the spectrograph is a blurred line and a distribution of intensity blended to right and left of the peak. This effect is shown in Illustration 4a′.

The radiation emitted by the atoms moving toward an observer has a greater frequency; the light emitted by the atoms moving away from him has a lesser frequency. The spectral line he sees therefore appears broader than the theoretical ideal line: it is composed of several frequencies, close together but not coinciding with the frequency that would be emitted by the atoms if they were at rest. Illustration 4a″ shows this effect in an actual spectrograph. The atoms are in disorderly movement because the source is extremely hot—a voltaic arc (the photograph is greatly enlarged).

By measuring the width of such spectral lines, it is possible to deduce the temperature of stars. This has been done for many years and astronomers have become familiar with the temperatures of stellar atmospheres.

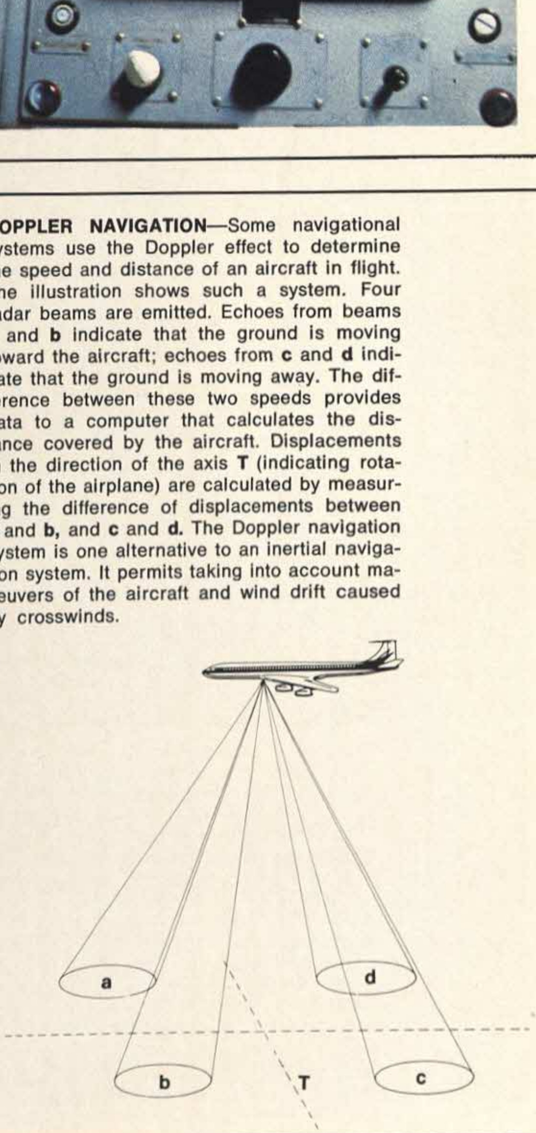
The Doppler effect also makes it possible for an observer to determine the thermal structure of a source. In the case of a voltaic arc, the central part is much hotter than the periphery. An example of this is shown in Illustration 4b. The spectral lines emitted by the central part have a wide distribution diagram. The peripheral part emits thinner lines because the plasma of the arc is cooler. A well-known law of physics states that a gas can emit and absorb the same radiation; the peripheral plasma in Illustration 4b, therefore, absorbs the radiation of the central plasma in the range corresponding to the central diagram. The resulting intensity distribution is shown on the right of the illustration.

Illustration 4b′ is a photograph (also greatly enlarged) of the reabsorbed spectral line of the voltaic arc. A spectral line having this kind of distribution intensity is said to be inverted. Structures with a much larger number of thermal shells can be analyzed by this method.

Stars rotate about a central axis. Various points of the stars' surfaces have speeds relative to an observer. The Doppler effect is used to measure this phenomenon and thereby determine the speed of rotation.

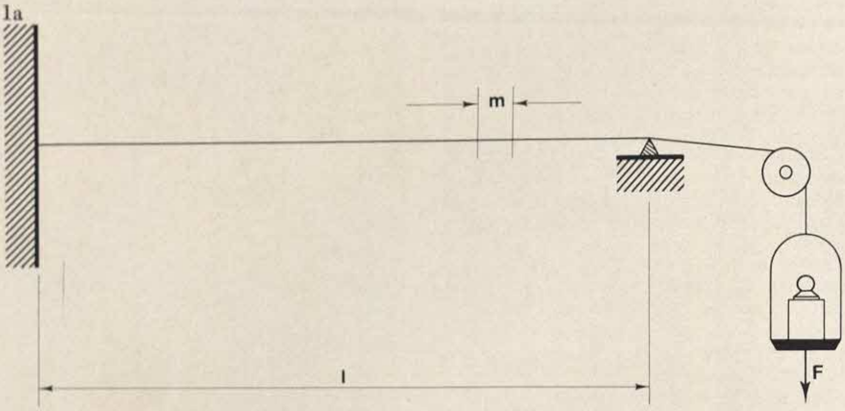
DOPPLER NAVIGATION—Some navigational systems use the Doppler effect to determine the speed and distance of an aircraft in flight. The illustration shows such a system. Four radar beams are emitted. Echoes from beams **a** and **b** indicate that the ground is moving toward the aircraft; echoes from **c** and **d** indicate that the ground is moving away. The difference between these two speeds provides data to a computer that calculates the distance covered by the aircraft. Displacements in the direction of the axis **T** (indicating rotation of the airplane) are calculated by measuring the difference of displacements between **a** and **b**, and **c** and **d**. The Doppler navigation system is one alternative to an inertial navigation system. It permits taking into account maneuvers of the aircraft and wind drift caused by crosswinds.

6

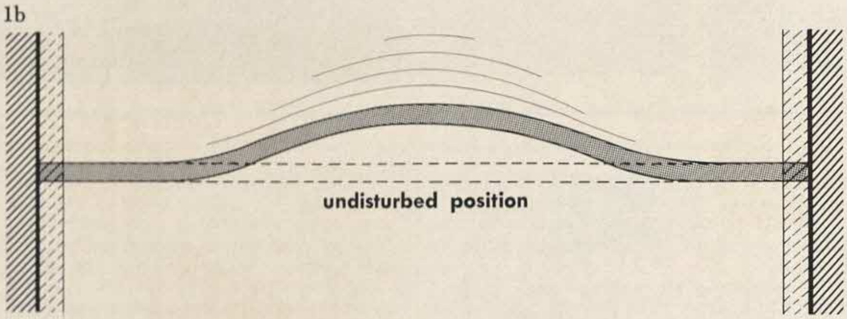


MUSICAL ACOUSTICS | air in rapid motion

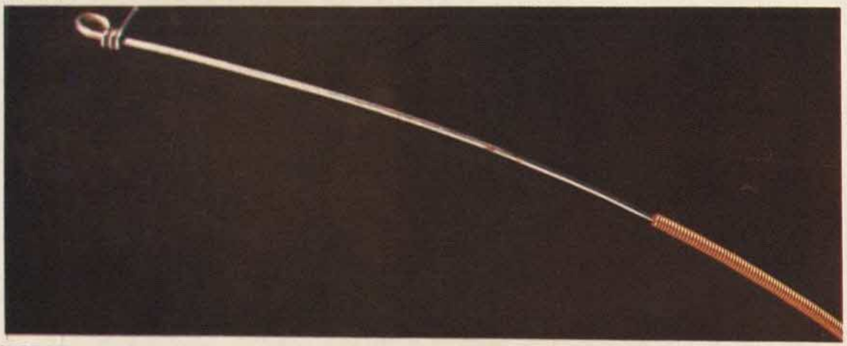
1a



1b



1c



VIBRATING STRINGS—As far back as two thousand years ago, Pythagoras discovered that the musical note emitted by a vibrating string is closely related both to the length of the string and to the tension under which it is strung. Provided that the mechanical characteristics of a string are known, a simple mathematical formula gives the frequency with which the string will vibrate under the influence of an external source of energy. In practice, however, the frequency of vibration is quite different from the fundamental frequency.

Illustration 1a shows the characteristics that must be known in order to analyze the vibration of a string. l is the length of the string, m is the mass of the string per unit length, and F is the force, or tension, exerted on the string. The following conditions must exist: the supports to which the string is attached at both ends are perfectly rigid; the string itself is perfectly flexible and its mass per unit length is uniform; the tension does not increase when the string is vibrating, and the air through which the sound is transmitted does not damp the string's vibration. Under these conditions, the fundamental frequency is given by the following formula:

$$f_0 = \frac{1}{2l}\sqrt{\frac{F}{m}}.$$

In practice, however, phenomena arise that interfere considerably with the theoretical conditions of vibration. The end supports are not perfectly rigid, and they do absorb a little of the vibration energy. The air in which the string vibrates has a slight damping effect. As the string is fixed at the ends, a part of its length (Illustration 1b) is unable to follow a vibrating motion, which diminishes the amplitude of the arc it can theoretically form. The mass per unit length of the string is not always constant over the total length because of very small variations in diameter. Finally, as the string moves from its rest position, the tension in the string increases. Some of these interfering effects may be taken into account by introducing a correction factor into the frequency formula: $f = f_0(1 + G)n$, where n is the harmonic of the vibration, and G is the factor taking the frequency variation into account. The value of this formula depends on calculating the value of G from the mechanical characteristics of the system. For example, the following formula takes into account the fact that the vibrating string is not perfectly flexible but slightly rigid; d is the diameter of the string and Y the coefficient of elasticity of the material of which it is made:

$$G = \frac{n^2\pi^3 d^4 Y}{128 l^2 F}.$$

The strings of some stringed instruments are usually made of gut, which is much more elastic than the steel wire of which piano, guitar, and mandolin strings are made. Nevertheless, in both types of strings actual frequencies are quite different from theoretical ones. Special tables take into account the different coefficients of elasticity of the two materials; these tables are used in the manufacture of stringed instruments. The correction constant, G, takes on yet another value when the vibration transmitted to the supports is considered, and yet a third value when the varying density along the length of the string is considered. The formula giving the value of G as a result of these other disturbing effects is more complex than the formula given.

An overtight string introduces an excess number of overtones, resulting in an unpleasant sound that is different from that of a mathematically ideal string. Such an unpleasant sound would approach that of "white noise," an inharmonious sound containing all audible frequencies that is usually produced electronically. For this reason, strings are made of very elastic material. In order to emit a sound, a string must do more than vibrate; it must also transmit sufficient energy to the surrounding air. The string must therefore be of certain dimensions. For example, strings emitting low frequencies transmit relatively little energy to the air. To emit a sound having a fairly high intensity, such strings must be long. Because their inertia must be high for them to vibrate slowly, such strings must have a fairly large diameter, which, in turn, implies a large mass per unit length, with the consequent danger of the strings becoming too rigid. The bass strings of a piano overcome this obstacle by their design. The string (Illustration 1c) consists of a thin steel core covered with a spiral winding of high-density copper wire that gives the string a high mass per unit length and a very high surface area in contact with the air into which the vibration energy is dissipated.

CYLINDRICAL AND CONICAL AIR COLUMNS—Sounds may be generated by the vibration of columns of air contained in variously shaped pipes.

Illustration 2a shows a pipe with an air column in vibration. On the left, air is introduced by a source of energy (which may be the player's lung power in the case of a trumpet, or a motor-driven pump in the case of an organ). The air is forced through a small orifice from which it emerges with a hissing sound similar to that of white noise. Because all frequencies are present in white noise, the design of the pipe selects the desired frequency at which the air inside the pipe is to be set in forced vibration, or resonance. In a pipe such as the one illustrated (closed at one end and open at the other), resonance takes place at a frequency given by the formula

$$f = \frac{nv}{4l}$$

Illustration 2a shows the points between which the length, l, of the pipe is calculated: v is the velocity of the sound wave in the pipe; n is the type of vibration. When $n = 1$, the fundamental frequency results; when n is equal to an odd integer (3, 5, 7, . . .), higher overtones result.

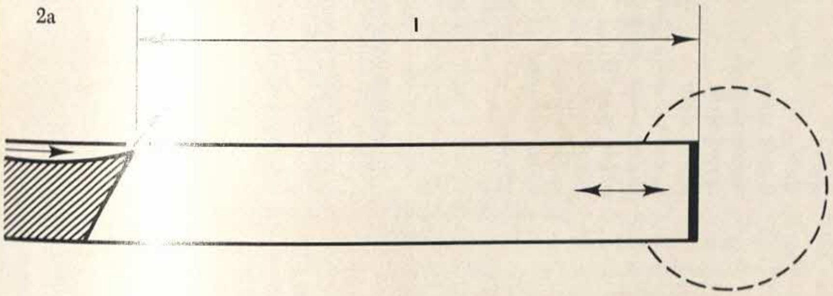
This formula becomes more complex when actual conditions that disturb the pipe's behavior are considered. The first such condition is that air motion within the cylindrical column sets the outside air in motion. This decreases the action of compression and rarefaction of the air (indicated by the double-headed arrow) that takes place at the closed end of the pipe, and thus alters the frequency of the vibrating air column. This disturbing effect cannot be calculated; only after experiments on several pipes have been carried out is it possible to determine a correction factor that can be applied to pipes similar in shape to the one used in the experiment; thus:

$$f = \frac{nv}{4(l + x)}$$

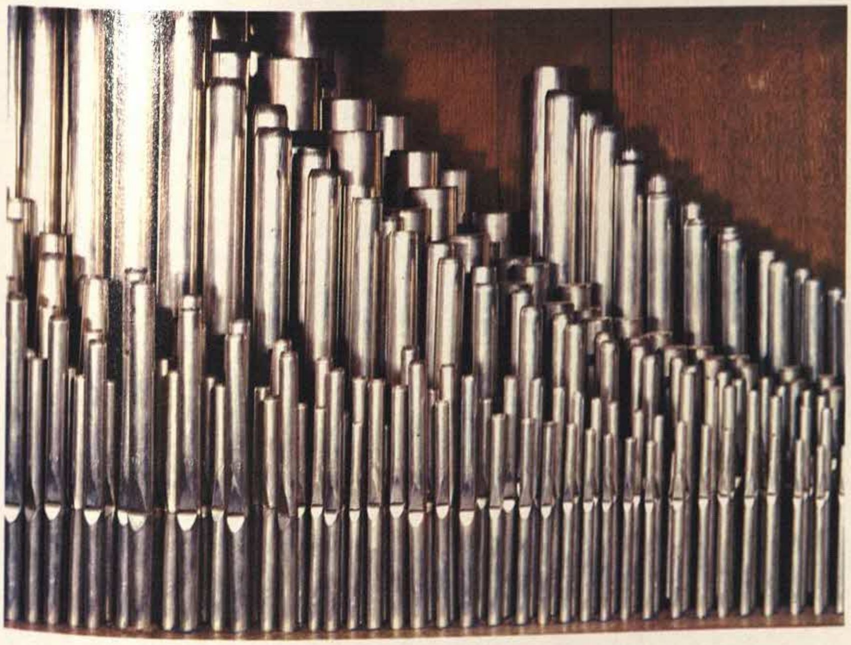
where x is the length that must be added to the original length of the pipe in order to produce the frequency f. An organ (Illustration 2b) works on the principle of the closed-end pipe. The tone quality of this instrument depends partially on the ratio between the length and diameter of its pipes.

A slightly different case is that of instruments where resonance takes place not in a cylinder, as in the case of the organ, but in an open cone of appropriate shape. Certain wind instruments are designed in this manner (although others have cylindrical resonators). Here, the frequency must be corrected by a factor that depends on the opening of the conical tube as in the case of the saxophone (Illustration 2c).

2a



2b

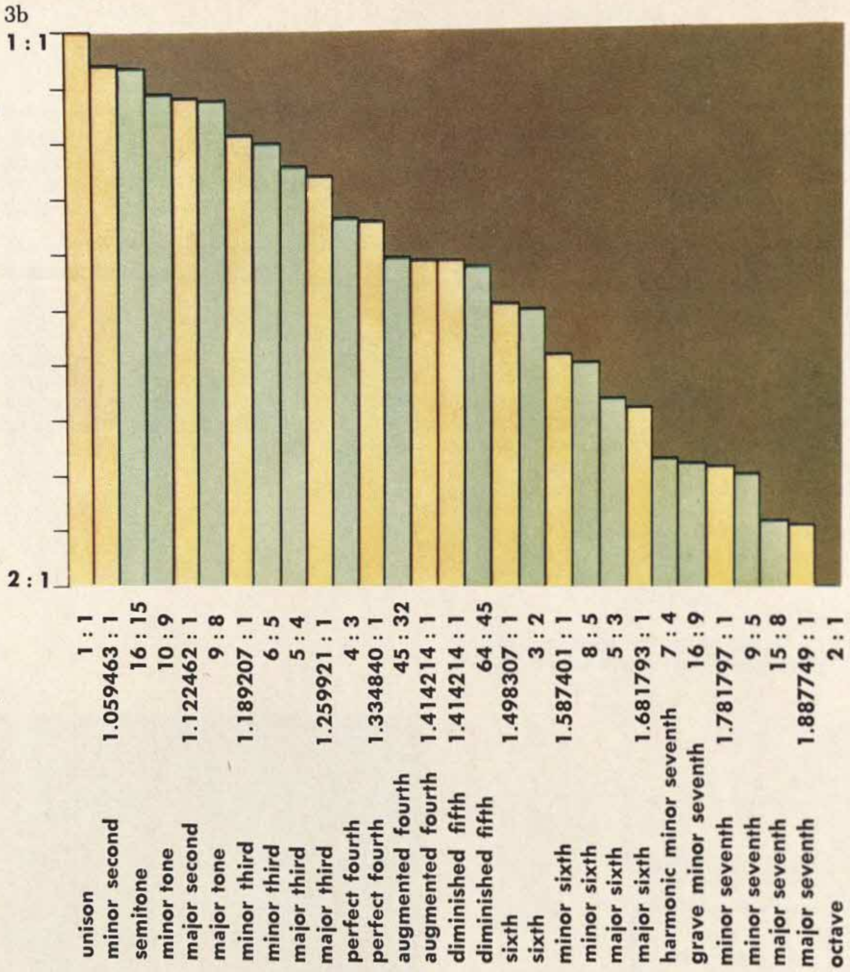


2c

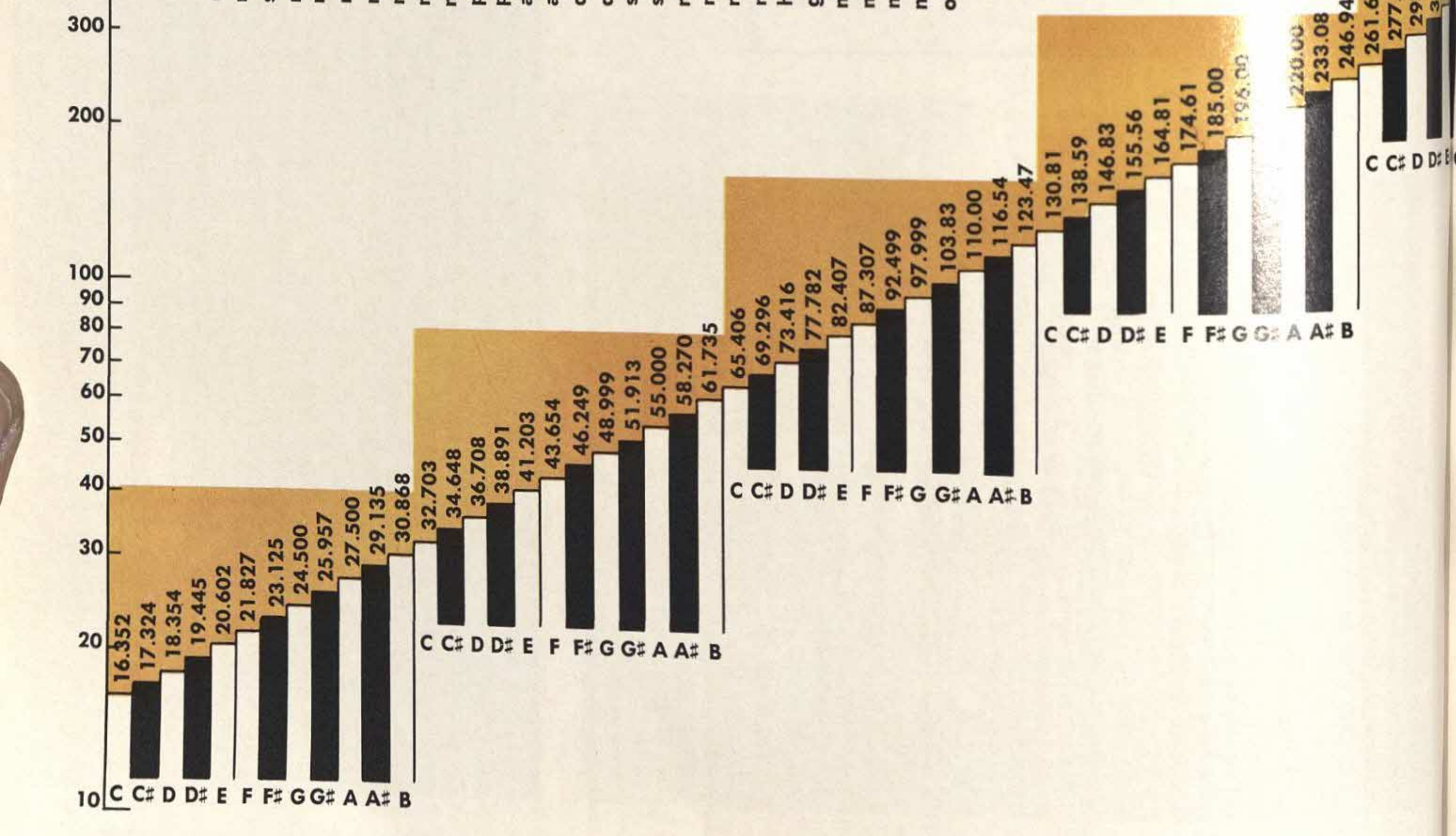
MUSICAL SCALES—In music, the ratio between two notes whose frequency differs by one whole multiple unit is called an octave; that is, eight notes (including the end ones) with a limit of seven intervals. There are other notes placed at suitable intervals with respect to these notes that divide the interval into unequal parts. The frequency of these notes is in a ratio to that at one end and is denoted by small whole numbers. Illustration 3a is a graphical representation of the position of the notes used in the common musical scale. The frequency of every note is given. The lowest, C_0, has a frequency of 16.352 Hz, while the highest has a frequency of 7,902 Hz. At frequencies below C_0 most ears can perceive no sound, whereas at frequencies above 7,902 Hz, only an unpleasant noise is audible.

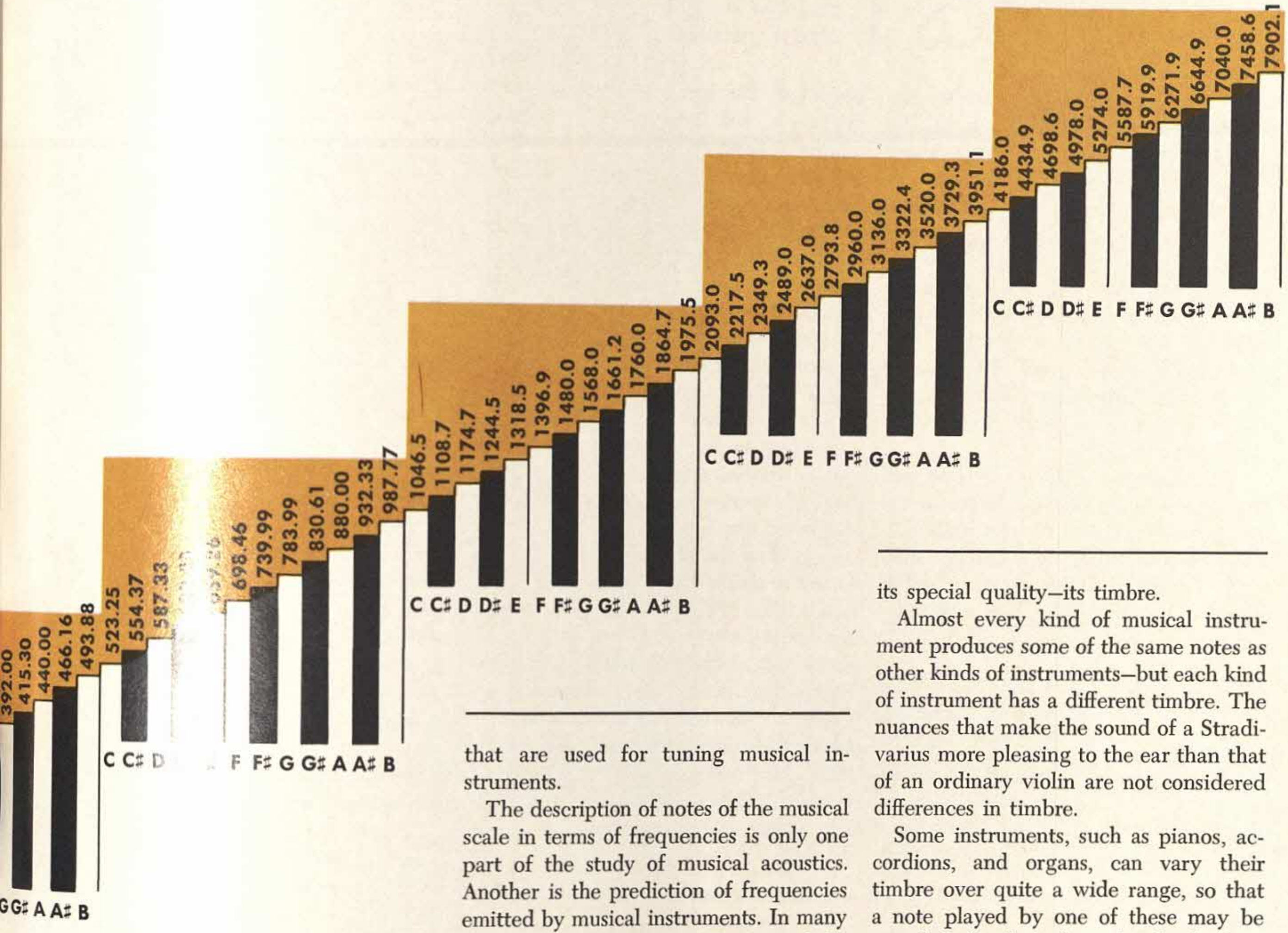
The ratios between the note frequencies (Illustration 3b) have names that depend on the numerical value of the ratio. For example, two notes having the same pitch are in unison; if their ratio is 2:1, the notes are an octave apart.

3b



3a





Musical acoustics is a branch of physics dealing with the generation of sound by musical instruments. A sound may be described as a vibration of the air—a vibration transmitted to the ears. Musical sounds are an assembly of sounds that are pleasing to the ear.

The most important characteristic of a musical sound is its fundamental frequency, which is the lowest note that the sound contains. The fundamental frequency, or, more simply, the fundamental, of a sound is indicated by a number. For instance, it may be said that a musical sound has a frequency of 440 hertz, abbreviated Hz, which means that the air vibrates at a rate of 440 times per second.

A musician does not often speak of a note in terms of hertz; instead, he calls the sound A440, knowing that it is the international standard of calibration of tuning forks and other standards of pitch that are used for tuning musical instruments.

The description of notes of the musical scale in terms of frequencies is only one part of the study of musical acoustics. Another is the prediction of frequencies emitted by musical instruments. In many cases these frequencies are emitted by vibrating strings, as in the piano, harpsichord, and violin, or, in other cases, by vibrating air columns, as in the pipe organ and other wind instruments. This article examines the phenomenon of sound emission and, in particular, why the timbre, or sound quality, of a given note played on a trumpet differs from the timbre of the same note played on a violin.

TIMBRE OF MUSICAL INSTRUMENTS

Timbre results from the composition of the sound itself. For example, a vibrating piano string, if properly tuned, produces a particular sound that is identified as a specific musical note. This sound is made up of a fundamental and overtone frequencies, which have values corresponding to whole-number multiples of the fundamental. The intensities of these overtones are lower than that of the fundamental. The overtones modify the fundamental, thereby giving the sound its special quality—its timbre.

Almost every kind of musical instrument produces some of the same notes as other kinds of instruments—but each kind of instrument has a different timbre. The nuances that make the sound of a Stradivarius more pleasing to the ear than that of an ordinary violin are not considered differences in timbre.

Some instruments, such as pianos, accordions, and organs, can vary their timbre over quite a wide range, so that a note played by one of these may be mistaken for that of another kind of instrument. Most organs, for example, can imitate many different instruments, even stringed ones. Music written for an instrument having a single and well-defined timbre, such as a violin, has a particular character, and the composer may be concerned with getting his acoustic effects from this timbre alone. However, the composer's inspiration may occasionally induce him to make use of the effects obtained by varying the timbre of the violin. A violin string, for instance, changes timbre according to whether it is vibrated by bowing or by plucking. By means of electronics, magnetic tape, and loudspeakers, sounds have been produced that have any desired number of overtones, thus duplicating the timbre of any musical instrument or even creating new timbres that cannot be obtained from existing instruments. The timbres so produced may be either gradually or abruptly varied at will. Electronically produced timbres open up new horizons in musical acoustics—horizons that have been only partially explored.

ARCHITECTURAL ACOUSTICS

protection from the daily intrusions of noise

Architectural acoustics is chiefly concerned with two problems: the insulation of rooms from internal and external noises; and the elimination of sound reflection and resonance within the rooms themselves.

Noise within a room may have two different causes. First, a sound wave propagated in the air outside the room may strike the wall and set it vibrating, thus generating a sound wave that can be heard inside the rooms; second, machinery operating inside a building can cause wall vibrations.

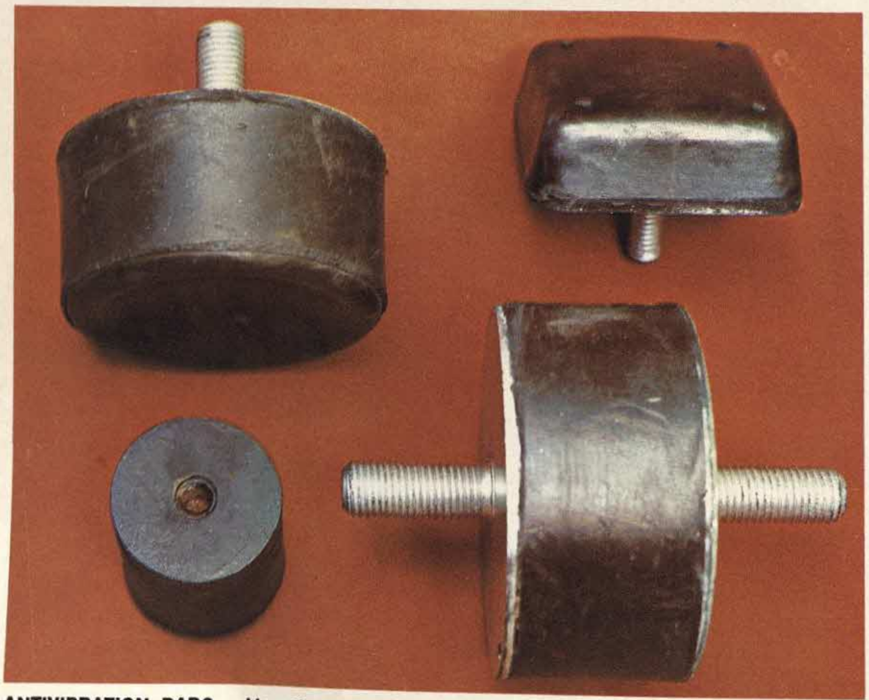
The problem of sound reflection and resonance is especially important in the design of film studios, concert halls, theaters, meeting halls, and similar structures. In these buildings, the proper reflection of sound is of paramount importance.

INSULATION OF NOISE SOURCES

Noise sources, such as engines and machinery, must be insulated so that the noises they produce will not be undesirably high. Bedding the machines on shock absorbers, like the ones shown in Illustration 1, will prevent the vibrations caused by the machines from being transmitted to the building itself. However, a building can also absorb vibrations that the machinery transmits through the air. Therefore, it is desirable to insulate the walls of a building so that they cannot absorb sound waves. This may be done by paneling the walls with materials that absorb sounds and will not transmit them to the walls.

Vibrations that have been absorbed by a building despite these measures will be propagated within the building and the energy of vibration will eventually be dissipated into heat energy. Rigid buildings, such as those constructed with reinforced concrete, transmit vibrations more easily than brick buildings. Every contact surface between brick and mortar reflects part of the vibrations, which are then more readily absorbed and dissipated. Because brick buildings generally have a greater mass for a given strength, they also absorb more vibrations. In a rigid structure, such as a steel ship, it is very difficult for the vibrations to become damped. The same is true of certain types of buildings. In buildings of concrete or steel, it is important to prevent vibrations from being propagated by the machinery internal to the structure.

1



ANTIVIBRATION PADS — Mounting noisy, vibrating machinery on these rubber shock absorbers prevents the vibrations from being transmitted to the building.

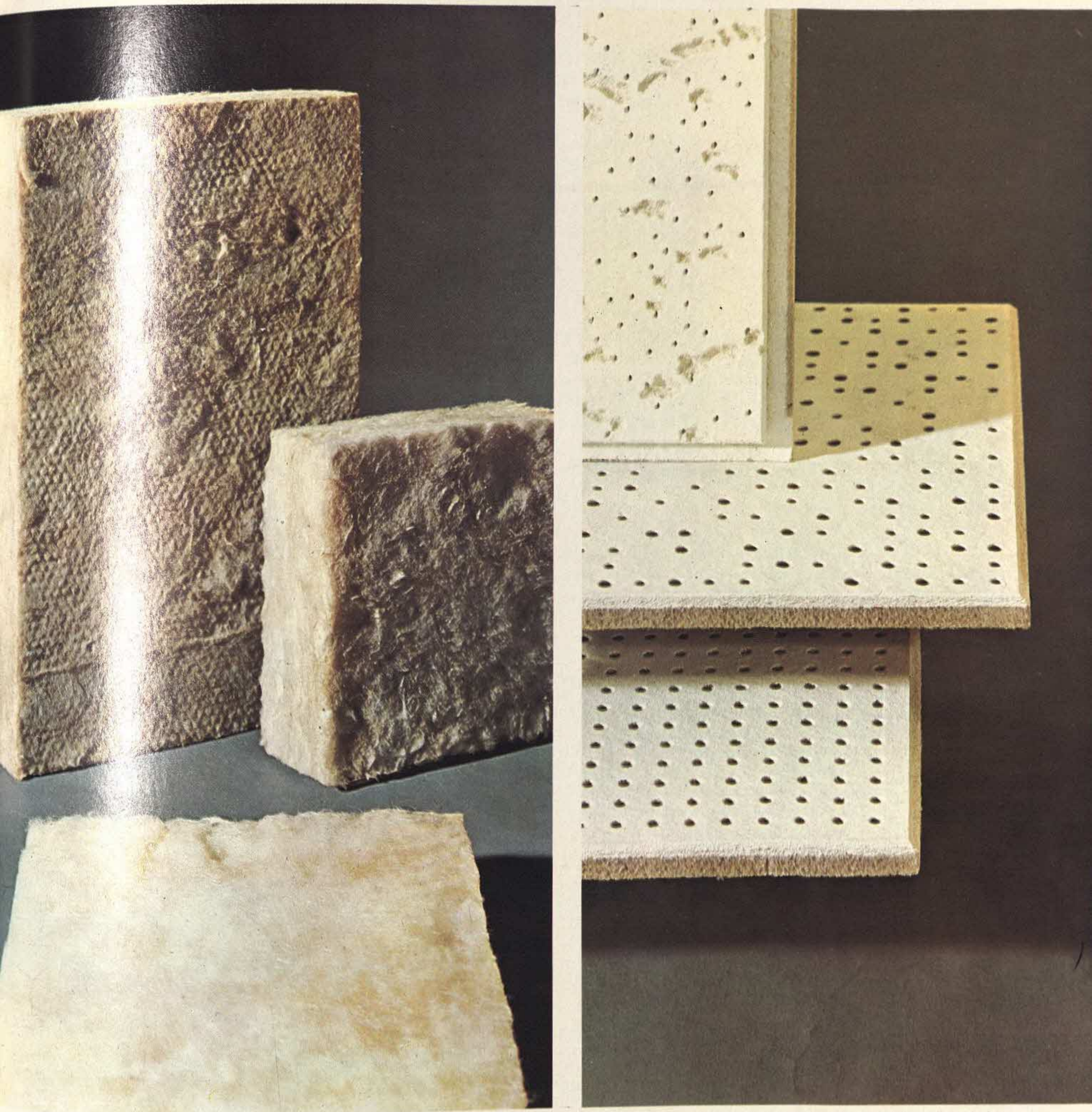
INSULATION OF EXTERNAL NOISES

Sound can enter a room in two ways: through openings in the walls, such as doors or windows; and through the walls themselves. When a sound wave strikes a wall, part of it will be reflected and part will be absorbed by the wall. That part of the sound wave absorbed by the wall will reach the inside of the room, although part of the transmitted sound wave will be absorbed by the wall. To insulate a room against outside noises when the mass of the walls cannot be increased, walls should be made of materials capable of absorbing much of the sound wave. The most frequently used materials for this purpose are: expanded polystyrene; glass wool; cellulose felt; perforated panels of rigid material covered with absorbent material; felts of rock wool, glass wool, and animal and vegetable fibers held together by a suitable binder; panels made of wood shavings and sawdust; and hollow panels with a flexible membrane. A sound wave that strikes such hollow panels will flex the fibers of the panel and compress the membrane without deforming the whole structure of the panel. Consequently, the panel will not communicate the pressure of the wave to the air on the other side. Some of these materials are so efficient in absorbing sound that they are used to insulate small open telephone booths in noisy locations.

Two panels (or two brick walls) separated by a distance are more efficient than a single panel or a single wall of twice the thickness. This is because the sound wave is reflected on the air-wall and wall-air surfaces between the panels or walls, and part of the sound wave is re-reflected into the first wall. However, it is essential that the two walls should not have a point of contact between them. In such a case, the vibrations would use such a point as a bridge from one wall to the other; whereas the air between two walls without such a point of contact would dampen the sound wave.

Doors and windows are more permeable to sound waves than the walls of a

2a 2b



INSULATING MATERIALS—These consist of special rock or glass wool (Illustration 2a) or rigid perforated materials covered with absorbent cotton (Illustration 2b). Such materials absorb sound waves and prevent them from being reflected, thereby eliminating resonance.

room. However, as with walls, the thicker doors and windows are, the better they insulate. A double door, or double glazing, insulates sound better than a single door or window panel.

The acoustic insulation of rooms not only prevents them from being disturbed by outside noises, but also protects them from noises generated within the rooms because the insulation also absorbs sound waves produced in the rooms.

RECOMMENDED SOUND LEVELS FOR DIFFERENT TYPES OF ROOMS

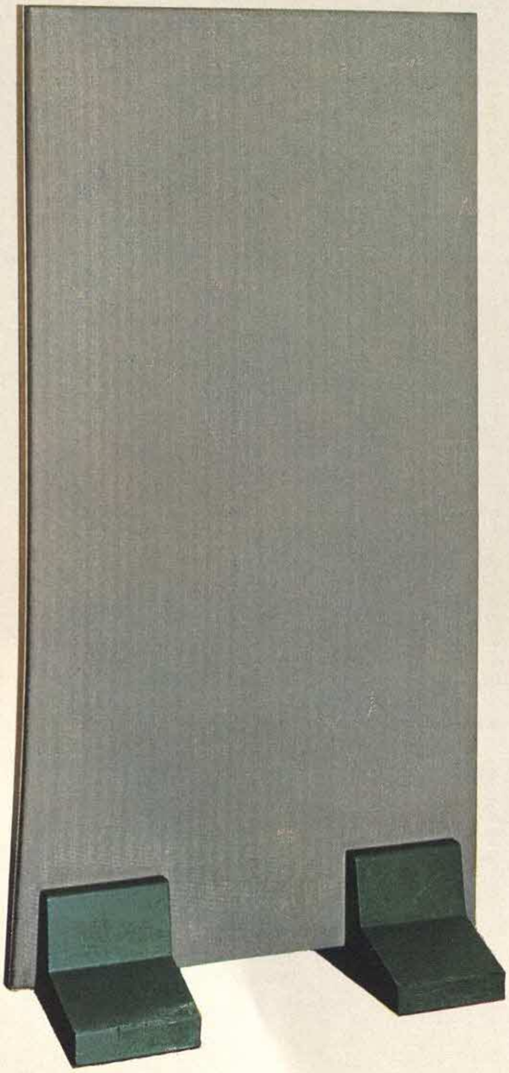
There are recommended sound levels for certain types of rooms, for example: 25–30 db (decibels, units of measurement of sound intensity) for radio and television studios and recording rooms; 35–40 db for hospitals, motion picture theaters, auditoriums, churches, hotels, homes, and classrooms; 40–45 db for small offices and conference rooms; 45–55 db for banks, department stores, and large offices; and 50–55 db for restaurants.

In a large office, where many typewriters and business machines are used simultaneously, the sound level would become unbearable unless the internal walls and ceiling were carefully lined with sound-absorbing panels.

THE ACOUSTIC PROPERTIES OF LARGE AUDITORIUMS: REVERBERATION TIME

When sound waves are emitted by a source situated inside a closed room, the waves generated by the source are propagated to the walls. At the walls they are partially reflected and partially absorbed. The part reflected continues to be propagated until it reaches another wall, at which point the process is repeated until the sound wave is completely absorbed. If the source continues to emit sound waves, the noise level in the room continually increases. The intensity of the sound inside the room will, therefore, increase for a certain period of time after the source has begun to emit sound waves and will stabilize itself at a certain level. If the source stops emitting sound waves, the intensity of the sound in the room will decrease until it reaches zero as the previously emitted sound waves are gradually damped.

3a

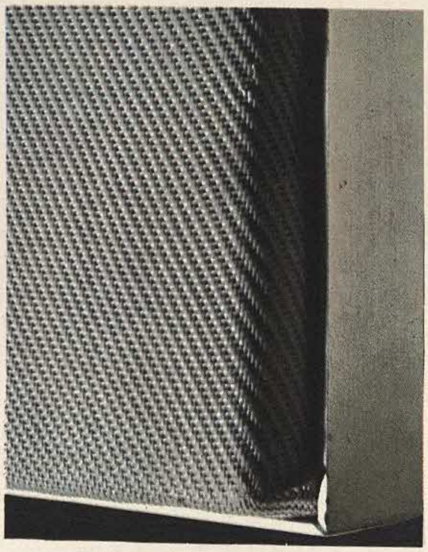


The damping is rapid if the room is small, because the sound waves will be reflected frequently. Strongly absorbent walls, even in a large room, will ensure rapid damping.

An extremely large room with walls that are relatively nonabsorbent will ensure quite slow damping. The time required for the sound intensity in a room to fall to a millionth of its original value is called the reverberation time and is designated by the Greek letter tau (τ). A room with zero reverberation time is one in which only the sounds emitted by the source are heard. There is no echo; such rooms are said to be echo-free.

The most suitable values of reverberation time vary with the use of the room. A reverberation time chosen to complement the types of sounds that will be heard within a room will add to the pleasure and ease of listening. Therefore, when designing a conference room, motion-picture theater, or concert hall, the geometric form and absorptive capacity of the walls must be so chosen that the room will have the desired reverberation time. A short reverberation time (less than one second) is essential for a conference room. A longer reverberation time would make it difficult to hear the voice of the speaker because his voice would be obscured by echoes. Some organ music, on the other hand, calls for a longer reverberation time, generally on the order of two seconds or more.

3b

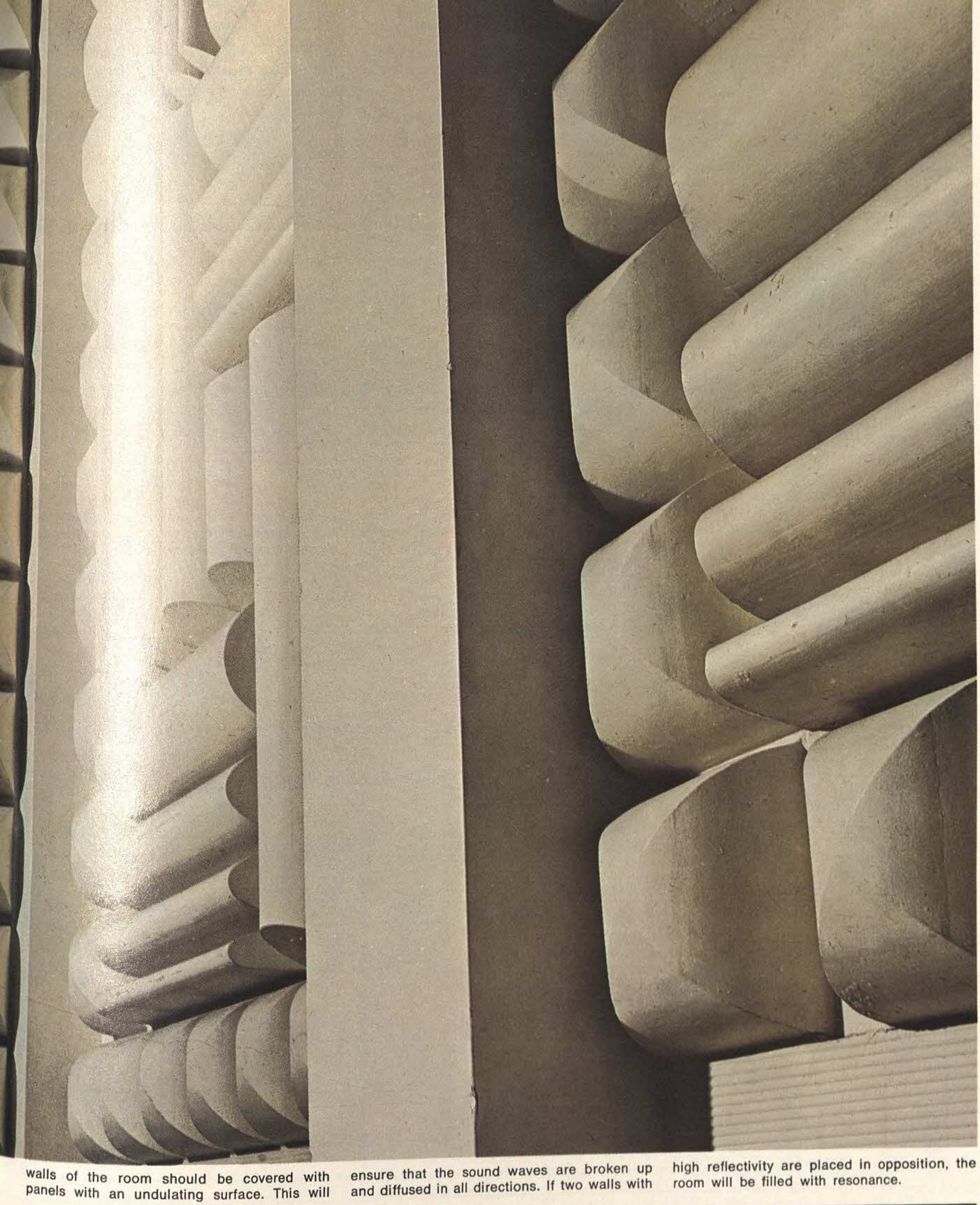


ISOLATION PANEL—This type of panel (Illustration 3a) is portable and can be adapted to almost any situation. The irregular surface of the panel shown in Illustration 3b is particularly suitable for absorbing and diffusing sound energy in a uniform manner.

A WALL FOR A RECORDING STUDIO—If the sound in a room is to be evenly diffused, the

DESIGNING THE ACOUSTIC CHARACTERISTICS OF ROOMS

Calculating the acoustic qualities of a room is a very difficult matter. If the form and dimensions of a room are known, the approximate acoustic characteristics of the walls (whether they will reflect,



walls of the room should be covered with panels with an undulating surface. This will ensure that the sound waves are broken up and diffused in all directions. If two walls with high reflectivity are placed in opposition, the room will be filled with resonance.

absorb, or diffuse the sounds) can be predicted. However, the best method of establishing acoustic characteristics is to experiment on a model of a section of the room. By means of such experiments, a correct value for the coefficient of absorption or diffusion for the walls and ceiling can be determined. Absorption is achieved by means of the same materials used in insulating a room against external noises. Sound is diffused, however, by modifying the surface of the walls. If the walls have an undulating surface, the sound waves will be broken up and diffused in all directions. It is important that two walls with high reflectivity should not be placed opposite each other, because such placement could lead to the establishment of stationary waves and the whole room could have a resonance similar to that occurring in a resonance box. Therefore, it is generally undesirable to construct a concert hall in the shape of a parallelepiped.

THE GEOMETRY OF FREEDOM

freedom and restraint translated into numbers

RESTRAINTS AND DEGREES OF FREEDOM

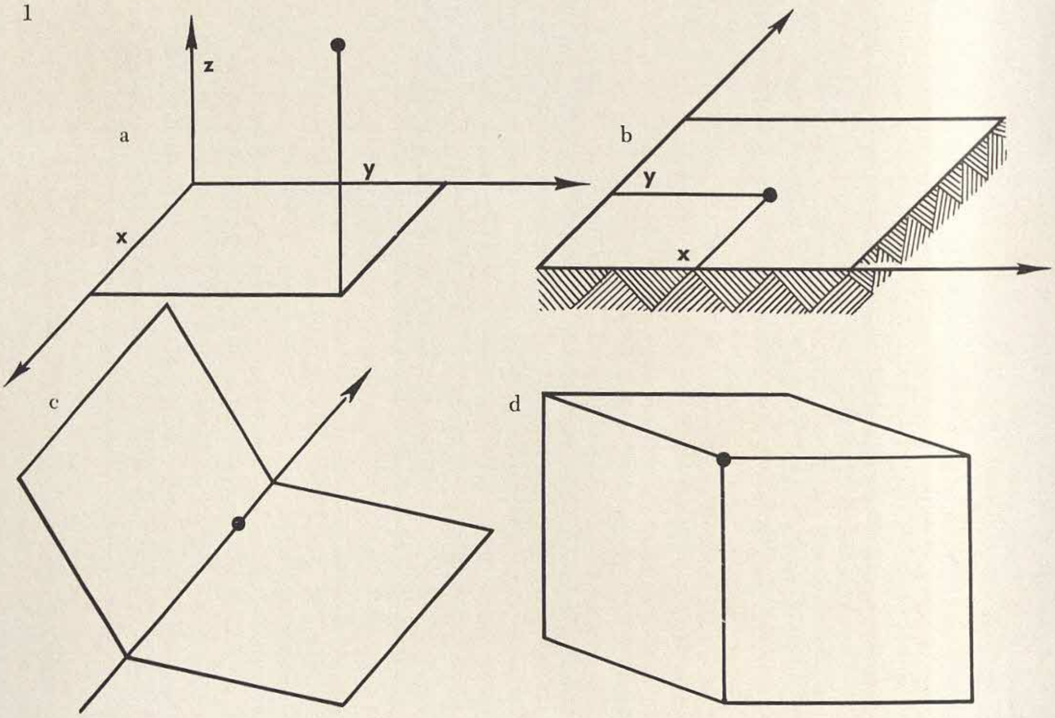
To know the exact position of a point on a straight line it is sufficient to know its distance from a point on the same line that has been chosen as the origin. Such a system has only one degree of freedom; that one coordinate (a number associated with the distance) is sufficient to fix the position of the point. "Degree of freedom" is the name given to the possibility of a point being displaced on one of the three coordinate dimensions. A point that is not subject to any restraints will, therefore, have three degrees of freedom. In other cases the point may lie in a certain plane, on a curved surface, or inside a closed box. Such a point is subject to restraints; that is, the possibility of its moving freely in space is diminished. A point is said to have as many degrees of freedom as the number of coordinates required to fix its position in space. Somewhat more complicated than a single point is a large body consisting of many points. Such a system, when left free in space, has six degrees of freedom (see Illustration 4).

A four-legged table is a large body and, therefore, possesses six degrees of freedom. If three of its four legs are placed on the ground, three points will be restrained and the table will thus be deprived of three degrees of freedom. In fact, the table is still left with the possibility of sliding on the ground, in either of two perpendicular directions, as well as the possibility of rotating. Three coordinates are, therefore, needed to fix its position. The length of the fourth leg must be such that its bottom will rest in the same plane as the other three—otherwise, the table will only rest on three legs

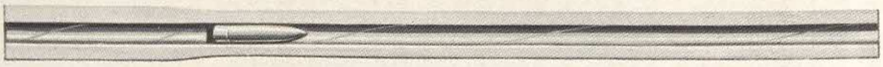
THE DEGREES OF FREEDOM OF A POINT—Illustration 1 shows how the degrees of freedom of a point are evaluated. Illustration 1a shows a point in space. Its position can be described by three coordinates: **x**, **y**, and **z**. Illustration 1b shows a point lying in a plane. One restraint has been imposed on the body, which has therefore lost one degree of freedom. As a result, only two numbers are needed to describe its position (one less than before). A point lying on a straight line—that is, a point at the intersection of two planes—is shown in Illustration 1c. One number is sufficient to describe its position on this line of intersection. Two restraints have been imposed on the body; it has lost two degrees of freedom. One number is therefore sufficient to describe its position. Illustration 1d shows a point obliged to lie at the intersection of three planes. It can only lie at the point common to all three planes; thus no number is needed to indicate where the point is to be found because the position of the restraint is known. Three restraints have been imposed; that is, the point has been deprived of the three degrees of freedom that it originally possessed.

These illustrations demonstrate the following rule: A free point possesses the maximum number of degrees of freedom and will lose one of them each time a restraint is imposed.

1



2



A BODY RESTRICTED TO A STRAIGHT LINE—Illustration 2 shows a bullet in the barrel of a rifle. The bullet is a body that can move freely in space, but inside the barrel it can move only forward or backward. In this case, the bullet is considered to be a body restrained by the barrel. The barrel constitutes a restraint that obliges the bullet to travel in a straight line. As long as the bullet is in the rifle, its position can be described by specifying the distance of its center of gravity from the breech of the gun.

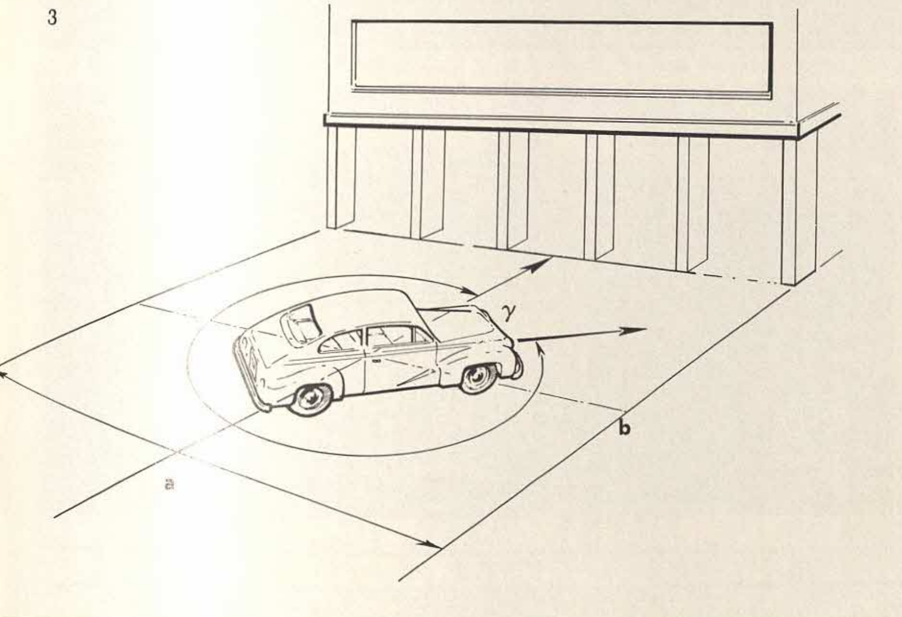
at a time, rather than on four. The addition of the fourth leg, therefore, does not stabilize the table, but rather makes possible a movement that was not possible on three legs.

The example given above does not involve a serious problem. However, in the case of a mechanism it may well be that it either will not function at all or will rapidly deteriorate. The design of a

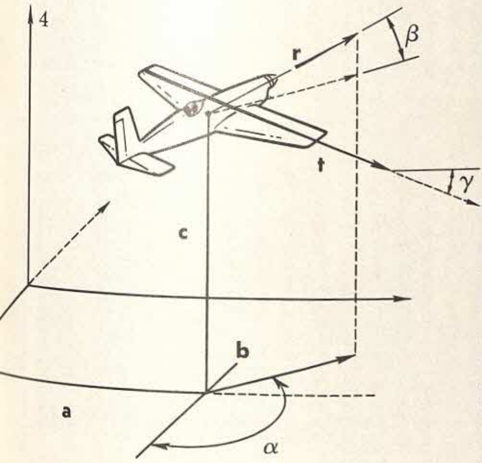
A BODY RESTRICTED TO A PLANE—This illustration shows an automobile in a large square. The automobile can move about in this square and can occupy any position. It cannot, however, descend below the level of the surface of the road, or rise above it. Hence, it can be said that the automobile is restricted to a plane.

To describe the position of the automobile mathematically, two perpendicular straight lines are drawn along the sides of the square and the perpendicular distance of the car from both of these lines is indicated. Two numbers will, therefore, describe its position. However, this is not sufficient. The automobile is not a material point, but rather a large body with a well-defined shape. If its four wheels are resting on the ground, its nose can be oriented in any direction with respect to the four cardinal points. To describe this orientation, a third number must be added that will represent the angle formed by the longitudinal axis of the automobile and one of the perpendicular lines that appear along the sides of the square. Therefore, three numbers are required.

3



4

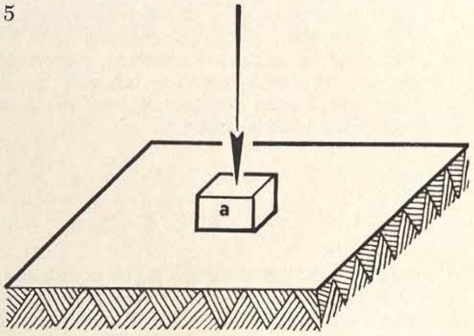


A BODY FREE IN SPACE—As long as an aircraft remains on the ground, its position can be described in exactly the same way as in the case of the automobile. If, however, the aircraft is in flight and performing aerobatics, as many as six numbers will be needed to describe its position and attitude. These numbers are used to fix the position of the aircraft in the following manner: Three numbers describe its position with respect to the ground. A perpendicular is drawn from the center of the aircraft to the ground. The point in which this perpendicular touches the ground represents the horizontal position of the aircraft. This position can be described by two numbers **a** and **b**, as in the case of the automobile. A third number **c** is needed to establish the altitude of the aircraft. If the aircraft were a geometric point, additional numbers would not be needed to describe its position. However, as an aircraft is not a geometric point, three additional numbers must be used to describe its orientation in space. As shown in the illustration, the aircraft is transformed into two arrows that intersect at its center; one of these arrows **r** is directed along the axis of the fuselage, the other **t** along the line of the wings. This latter arrow goes from left to right with respect to the aircraft's motion.

If the aircraft is performing aerobatics, the two arrows remain perpendicular to each other, but will have a random orientation in space. The orientation of this pair of perpendicular arrows is described by defining the following three angles: α is the angle formed by the arrow **r** with one of the horizontal coordinate axes; β is the angle formed by the arrow **r** with a horizontal plane passing through the center of the aircraft; and γ is the angle formed by the arrow **t** with the horizontal plane defined in the preceding angle.

IDEAL RESTRAINTS AND REAL RESTRAINTS—The surface shown here is an ideal restraint: a body can rest on it, but cannot pass beyond it. If the body is pressed against the surface with a certain force, the plane will react by resisting with an equal and opposite force that will stop the body. If the force with which the body is pressed against the restraint is increased, the surface will simply increase the reaction with which it opposes the attempt to pass beyond it. An ideal restraint will, therefore, resist the attempt to pass it—with a force that will be sufficient to counteract the one pressed against it, no matter how great this force may be. A real restraint, on the other hand, will eventually break. The top of a table constitutes a plane restraint. This restraint cannot be expected to prevent the downward passage of a body weighing, say, 2 tons; the table will collapse. In reasoning theoretical mechanics, ideal restraints are used. Should a restraint have to be constructed in a practical form, the maximum stress to which it will be subjected must be determined and the restraint constructed to be capable of resisting the maximum stress.

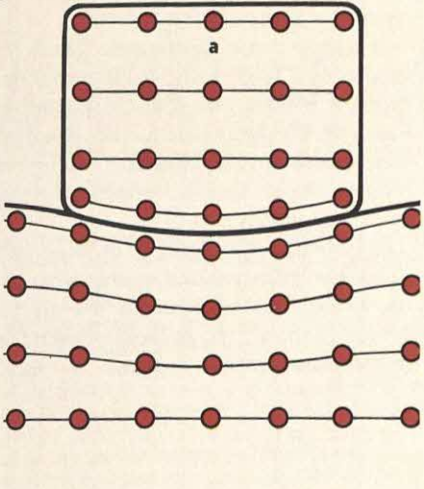
5



HOW THE REACTION OF THE RESTRAINT COMES INTO BEING—The parallelepiped is a solid and rests on a base. It is shown enlarged to illustrate what happens at the point at which an object is made to rest on its plane.

The parallelepiped consists of atoms that appear at regular intervals. If an attempt is made to move the atoms farther apart, or, alternately, to compress them, they will resist by exerting an opposing force. They behave as if they were held in place by springs. Therefore, if a body is placed against a steel plane, the thrust exerted against it will tend to compress the atoms, which will, therefore, press against the body just as much as it presses against them.

6



7

8

UNILATERAL RESTRAINTS—The surface of the Earth constitutes a restraint for aircraft in the sense that they cannot descend below it. However, the restraint is unilateral. It prevents an aircraft from descending below it, but it does not prevent the aircraft from rising to any desired height above it.

BILATERAL RESTRAINTS—A pair of rails constitutes a bilateral restraint because it prevents a train from becoming displaced either to the left or to the right of it. Unilateral and bilateral restraints can be easily classified when the geometric characteristics of the bodies that constitute the restraints are known.

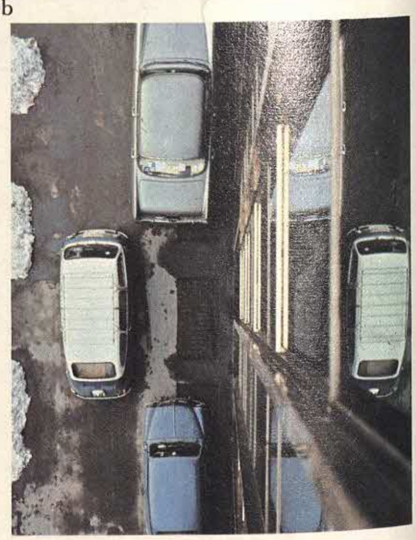
mechanism must necessarily take into account the degrees of freedom left to the moving parts. As mentioned above, there are six degrees of freedom for large bodies. The study of the degrees of freedom of large bodies has many applications, even outside the field of mechanical engineering. In physics, for example, such a study allows an investigation of how the energy that a molecule receives in the course of its collisions with other molecules is distributed inside it.

If a molecule of a diatomic gas is regarded as a large body, it has seven degrees of freedom—six that correspond to those of any large body, and one that corresponds to the vibrating motion of the two atoms moving toward and away from each other.

A system that is made up of three particles has nine degrees of freedom, because three different coordinates are required to specify the position of each one of the three particles in space.

9a

9b



"HOLONOMOUS" AND ANALONOMOUS RESTRAINTS—The plane of the floor of a room constitutes a "holonomous" restraint for a chair (Illustration 9a). This means that although the chair is in contact with the floor it can still be dragged to any point in the room, on any trajectory.

An automobile resting on a plane can also reach any desired point (Illustration 9b), but it cannot always go there by the shortest available route. Its steering system permits it to follow only certain trajectories (the automobile cannot just slide over the ground). The automobile shown at the center of this illustration cannot easily enter the parking space at its side, even though this space is sufficiently large to accommodate it, because the type of restraint to which the car is subject would require it to carry out an excessive number of maneuvers in the process.

EQUILIBRIUM | an agreement among forces

In a general sense, a body is in a state of equilibrium when the resultant of forces acting on it is zero. All bodies possess a natural disposition for reaching a state of equilibrium: for example, a stone rolling down the side of a mountain will stop when it reaches a state of equilibrium. This principle is true for a variety of motions, such as the position a pendulum assumes when it is in a state of equilibrium and the reason for the position.

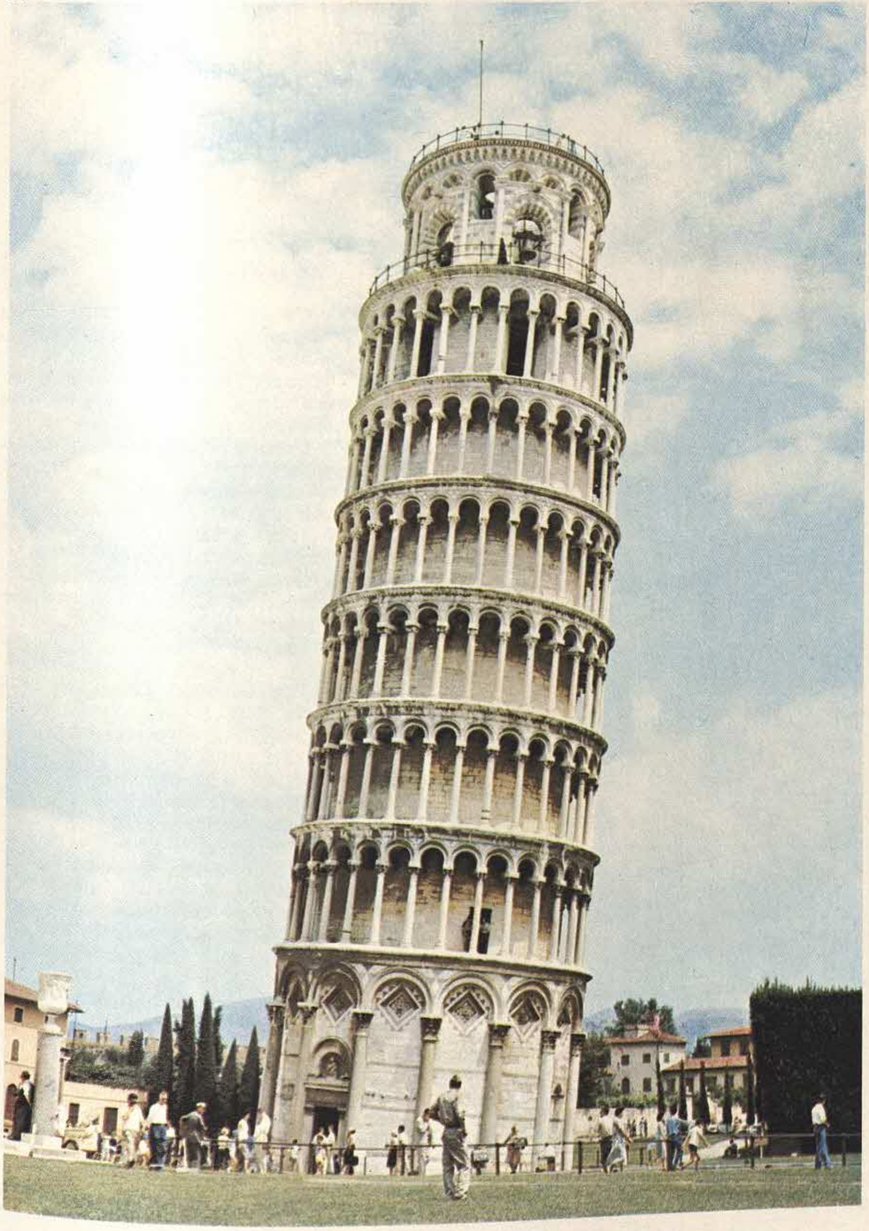
Further study of the conditions in which a state of equilibrium occurs enables engineers to design stable structures.

WHAT IS EQUILIBRIUM?

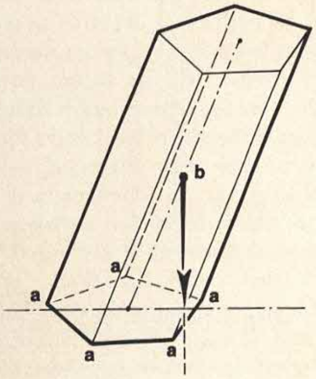
The first principle of mechanics states that (a) when a body is at rest or (b) when a body in motion is not subject to any forces it will (a) remain at rest or (b) continue to move at a constant veloc-

THE EQUILIBRIUM OF PRISMS—A prism can be in equilibrium only if its center of gravity **b** lies on a vertical line that passes through a point inside the base polygon of the prism, that is, the polygon **aaaaa** formed by the intersection of its lateral surfaces with the ground. In the case of the prism shown in Illustration 1a, the equilibrium subsists. An example of an equilibrium of this type is provided by the leaning tower of Pisa, which remains standing in spite of its pronounced inclination. It is possible to construct a solid of homogeneous material, on a rectangular base, that will remain standing in spite of projecting laterally beyond its base polygon by an infinite amount. The solid could be constructed of bricks or similar rectangular objects of equal size. The elements should be arranged horizontally, as indicated in Illustration 1c, which shows a side elevation. The first element is placed in any position; the second is placed on top of the first in such a manner that it projects by one quarter of its length. The third element is placed on top of the second in such a way that it projects by one fifth of its length beyond the second, and so forth. The subsequent elements will project by one sixth, one seventh, one eighth, and gradually less. Illustration 1c shows how the stack of elements continues to project laterally beyond its base, and, continuing to infinity, will eventually project by an infinite amount. In fact, the position of the center of each brick is displaced from the center of the first brick by the same amount as the sum of the displacements of each brick with respect to the one below it: 1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9 + . . . This succession of numbers constitutes a harmonic series whose sum is not a finite number. Therefore, as elements are added, these will always project more and more, without any limit. The interesting fact is that the center of gravity of this stack of objects, no matter how high and projecting it may become, will always remain on a vertical that passes through a point inside the first element: the stack will, therefore, remain standing in spite of its paradoxical form. The demonstration of this fact is not elementary, but it can be verified by summing the parallel forces consisting of the weights of the first and second elements, then the first two and the third, the first three and the fourth, and so forth.

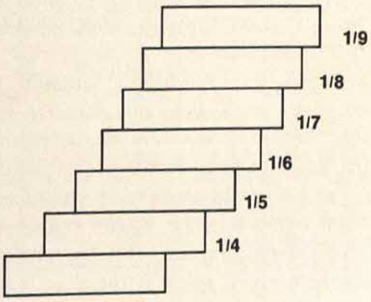
1b



1a



1c



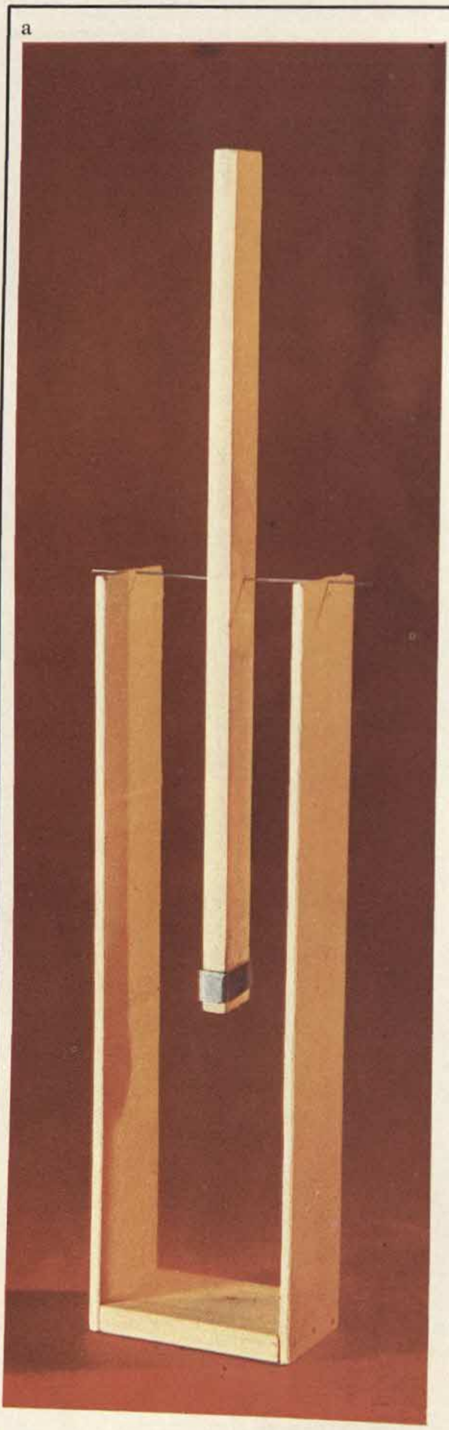
ity, or with a uniform motion in a straight line. Either of these conditions will prevail only if no forces are acting on the body, or if the resultant of forces acting on the body is zero. For example, a body suspended by a cable is subject to the force of its own weight and the pull of the cable, the latter of which is equal to the weight of the body. The two forces have equal intensity; the sum of their actions is thus zero and the body remains at rest (or, in a state of equilibrium). If a body is to remain at rest the forces acting on it must completely balance each other. However, if a body is subject to forces that cancel each other out, that is, equal and opposite forces, it does not necessarily mean that it will be stationary. For example, if a table tennis ball is dropped from a great height it will attain a constant speed after it has fallen a few feet. From this moment on, the weight of the ball and the force of the aerodynamic resistance that acts on it are equal and opposite; that is, they cancel each other.

Bodies are sometimes subject to various forces, the resultant of which is not that of a single force but of equal forces influencing the body in opposite directions, which causes it to rotate. To keep the body at rest, another couple of forces of equal intensity but directed in opposite directions must be applied. In the general sense, a body is in a state of equilibrium when the resultant of the forces that act on it is zero. For a more complete definition of the concept, however, other factors must be added.

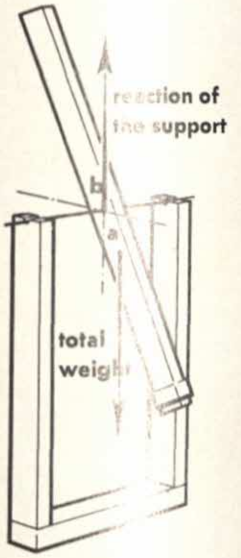
It must be established whether the forces that act on the body do so in such a direction as to reestablish the equilibrium if it were to be disturbed, or whether the forces would destroy it completely. A stick balanced on a juggler's nose is in a state of equilibrium, but the slightest disturbance will bring it toppling down. On the other hand, an object placed at the bottom of a receptacle will not fall out if the receptacle is slightly inclined.

2

a



b

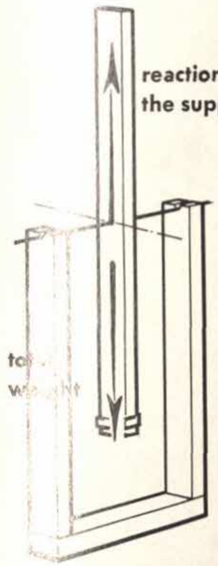


THE EQUILIBRIUM OF A PENDULUM—A study of rudimentary pendulums reveals a great deal about equilibrium.

Illustration 2a shows a pendulum that is formed by a piece of wood about 40 cm long, 3 cm wide, and 1 cm thick (about 16 × 1 × 0.5 in.). The center of this piece of wood has been pierced with a needle so that the ends of the needle project on both sides. At one end of the piece of wood is a lead weight. This rudimentary pendulum is suspended from two boards that are affixed vertically to a small wooden base. The needle serves as a hinge, about which the pendulum rotates, and is placed on the two vertical boards. The pendulum is now free to oscillate. However, its oscillations are not of interest here, but, rather, the position it assumes when it is in a state of equilibrium and the reason it assumes this particular position. As the illustration shows, the equilibrium position of the pendulum is the vertical one in which the lead weight comes to rest immediately below the suspension point (the needle).

If the pendulum is positioned so that it is inclined with respect to the vertical, as in Illustration 2b, it is subject to two forces: the total weight (made up of the weight of the piece of wood, and the lead weight), which is indicated by an arrow pointing vertically downward, and applied at point **a;** and the reaction of the support that sustains the pendulum and acts upward, indicated by an arrow pointing upward

c



and of the same length as the first arrow because the force of this reaction is equal to the weight of the body. This force is applied at the fulcrum **b**. These two forces constitute a couple; that is, they are two parallel, equal, and opposite forces, and their effect is that of causing the pendulum to return to the vertical position.

The equilibrium of the pendulum is, therefore, a stable one (as shown in Illustration 2c): when the pendulum is in the vertical position, the weight and the support reaction act along the same straight line and the couple has no effect. In fact, the two forces act in opposite directions along the same straight line and cancel each other.

When another slightly lighter weight is attached to the other end (Illustration 2d), the pendulum will still be in equilibrium in a vertical position. The equilibrium will still be a stable one, but less so than in the preceding case.

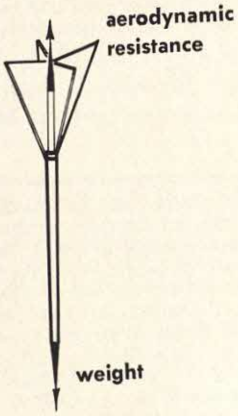
In fact, point **a** (at which the weight is applied) and point **b** (at which the support reaction is applied) are now very close to each other, and the couple they constitute is now far less effective in straightening the pendulum when it is placed in an inclined position. If the weights at either end of the center piece of wood were equal (or if neither end were weighted), the pendulum would come to rest at any position in which it was placed—a state of indifferent equilibrium.

d

3

a

b



EQUILIBRIUM IN UNIFORM MOTION—A body resembling a dart has been constructed to aid in studying the equilibrium of the forces acting on a body in motion. This body consists of a piece of drinking straw about 5 cm (about 2 in.) long, as shown in Illustration 3a. A ball of lead shot is glued inside one end of the straw, making the point much heavier than the rest of the straw. At the other end of the straw are four fins constructed of pieces of paper and glued to the straw. When the dart is dropped from a height of a few feet, the speed of its descent will become uniform after it has fallen a short distance. The interesting thing to observe in this experiment is that the dart, no matter in what position it is dropped, will fall only a very short distance before it assumes a vertical position with the tail uppermost, as shown in Illustration 3b. This illustration points up the analogy between this system and the pendulum discussed in the preceding illustration. The weight of the object is concentrated in the point, as represented in the figure by the arrow pointing vertically downward. On the other hand, the aerodynamic resistance acting on the dart is a force directed opposite to that of the motion, as indicated by the arrow pointing upward at the center of the tail. When the speed of the dart has become uniform, the two forces acting on it are perfectly aligned and equal, and the dart thus assumes a perfectly vertical position (as in the case of the pendulum subject to its own weight and the reaction of the support). This experiment illustrates positional equilibrium in a uniform rectilinear motion. In an analogous manner, the stability of an aircraft depends on the equilibrium between the forces (thrust of the engines, air resistance, weight) that act on the aircraft by virtue of its motion and its particular mechanical characteristics.

THE COMPOSITION OF FORCES

Although an immobile, or static, object—such as a bridge, a building, or even a man—is said to be at rest, the object is being acted on by a variety of forces. Because these forces cancel out each other, the object is in a state of equilibrium. This state of equilibrium was not attained by chance. It was worked out in advance by the object's maker through the application of statics, a branch of mechanics that deals with bodies at rest.

It is important to know the conditions under which a body does not move. All static, or nonmoving, structures, such as bridges, buildings, and tunnels, are subject to the forces of their own weight, the weight force of the objects they support, the force of wind and sometimes of water, the pressure of the Earth, and forces of expansion or contraction caused by heat or cold. Obviously, the structures must remain at rest, regardless of the nature or number of forces involved. A poorly designed tower may lean in the wind and fall; a building with a poorly designed foundation may be unsafe, as one example, because its foundation has been undermined by water. The design engineer must anticipate all the forces that will act on his structure and see that they are equalized so that the completed structure will be stable.

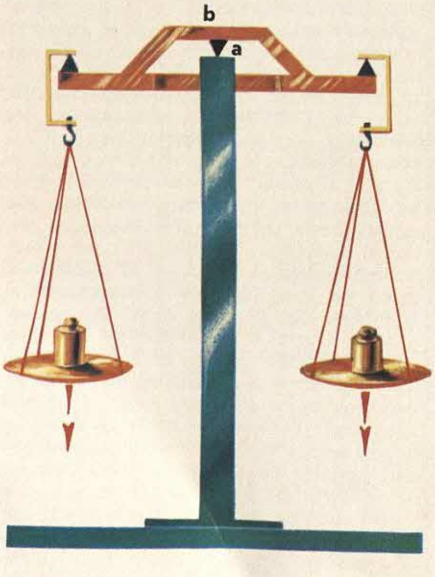
Two ways exist to equilibrate the forces on a structure. Equilibrium can be created through active forces, such as weights, or through passive forces, such as restraints. From experience, the designer knows the most important tools of structural engineering.

1

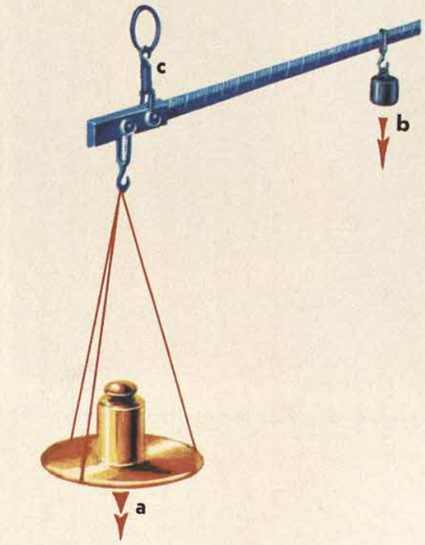
THE SUM OF PARALLEL FORCES—The forces acting on the scales of a balance are represented by arrows. In the study of statics, it is important to know the sum of the forces that are represented by arrows pointing in the same direction—or in all directions. In Illustration 1a, equal weights have been placed on both scales. Each of these weights is represented by an arrow pointing downward; if the balance is equilibrated, the weights are equal. What force is acting on the central knife-edge **a,** which supports the yoke of the balance **b**? Ignoring the weight of the mechanical parts of the balance, it is obvious that the weight acting on the support of the yoke is equal to the sum of the two weights on the scales. The resultant force at **a** is directed downward and its point of application is at the center of the yoke. Such forces are called parallel because they act in the same direction.

In a slightly more complicated example (Illustration 1b) the two forces that make up the resultant have different values although they are still parallel. The forces are applied to the arm of a kind of balance known as a steelyard. The heavier weight is resting on the scale **a** and the lighter weight **b** is movable. How much weight is acting on the suspension hook **c**? As in the previous example, the weight at **c** is equal to the sum of the two weights **a** and **b.** What is the position of the suspension hook **c** in relation to the hook supporting the movable weight **a**? It divides the distance between the point of application of the scale and that of the movable weight into two parts inversely proportional to the amount of the two weights **a** and **b.** Weight **b,** which exerts the smallest amount of downward force, is farther from the suspension hook than is weight **a.**

a



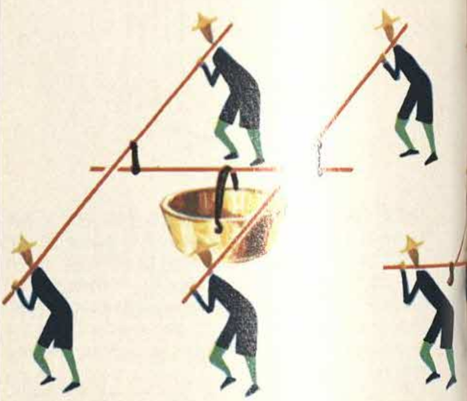
b



2

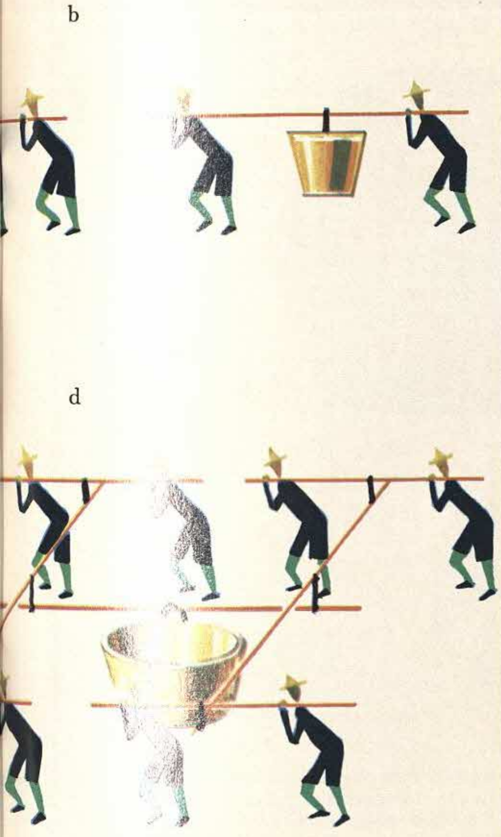
a

c



THE CHINESE WATER CARRIERS—In ancient China, water was carried in large buckets. When filled, these buckets were so heavy that several men were required to lift one of them. The Chinese invented a system that allowed the buckets to be carried so that each man carried a fraction of the total weight equal to the weight of the bucket divided by the number of men. In this way, no one could avoid bearing his fair share of the load. A simple method of transportation (Illustration 2a) is to suspend the bucket from the middle of a pole and place a row of men on either end. This method has one serious defect: if a man stoops slightly, he can pretend to be bearing his share of the weight, leaving the burden to his less artful colleagues.

In the first stage of the development of the Chinese method (Illustration 2b), the bucket is suspended from the middle of the pole and its weight sustained by one man at each end. Since each of the two men is the same distance from the bucket and the bucket is hung from the exact middle of the pole, each man bears exactly half its weight. Neither of the

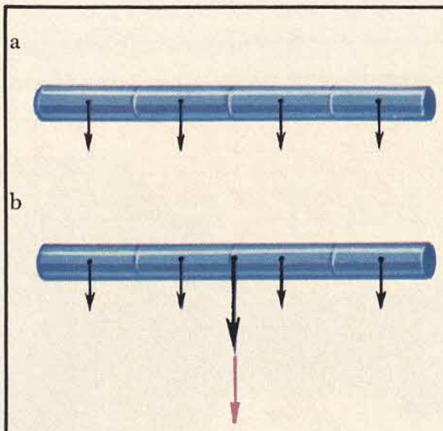


men can avoid bearing his half of the load because if he stoops, the bucket goes down with him. If the bucket is too heavy for two men, four men can be used.

In the final stage of the Chinese system (Illustration 2c), the pole supporting the weight of the bucket is supported at each end by a transverse pole perpendicular to the first. A flexible joint attaches each of the transverse poles to the first suspension pole. One man is placed at each end of each transverse pole. The weight of the bucket is divided in two at each end of the primary suspension pole. Each half of the weight is halved again at the ends of each of the transverse poles. Here, as in Illustration 2b, it is not possible for a man to shirk his share of the weight. If he stoops too far, the load will touch on the ground, and his deceit or exhaustion will be apparent to his colleagues.

The Chinese system can be adapted to distribute the weight among any even number of men. Illustration 2d shows how the weight is distributed among eight carriers.

3

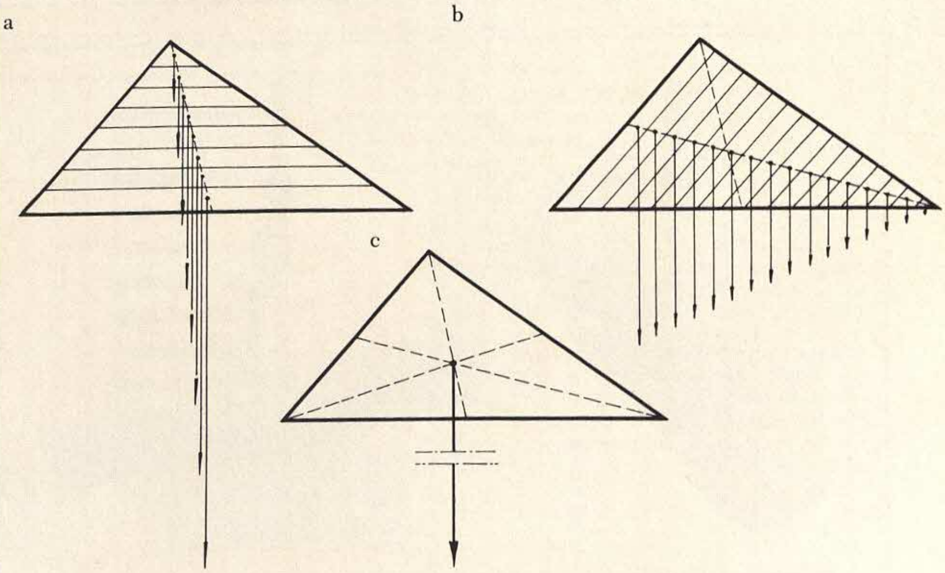


THE COMPOSITION OF PARALLEL FORCES—This iron bar (illustration 3a) is perfectly cylindrical in form. Imagine that it is divided into segments of the same length. Each segment will have the same weight. The weight of the elements is represented by arrows of equal length. The weight of the whole bar is evidently equal to the sum of the weights of the elements. To find this sum, add two by two the forces that are equidistant from the ends of the bar. The result is that a number of segments accumulate in the center of the bar equal to half the number of fractions of the bar. The arrows in the center (Illustration 2b) are twice the length of those representing each part. The sum of the segments is represented by the two-color arrow at the center of the bar. The weight of the bar is the resultant force due to the weights of all the parts. The weight acts through the center of the bar. The point through which the weight acts is called the *center of gravity.*

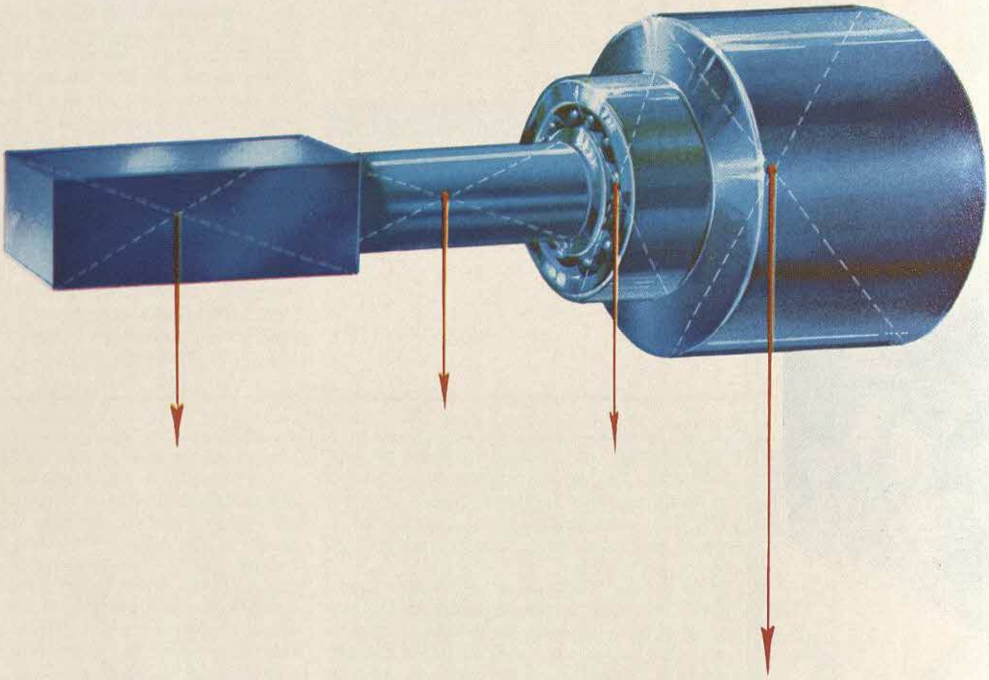
4

THE CENTER OF GRAVITY OF A PLANE SURFACE—Suppose a triangle of any shape has been cut from a sheet of paper. Like the metal bar, it can be thought of as made up of several elements, each having a weight applied to its center.

How is the resultant of all these forces found? Where is its point of application? The first step is to simplify the problem. The triangle (Illustration 4a) is divided into a number of parallel strips, all of the same width. The center of gravity of each strip will be at its center, just as in the iron bar of Illustration 3. It does not matter that the angles at the opposite ends of the strips are different. If the strips are very narrow, their centers of gravity will fall exactly in the middle of the strips, or at a negligible distance from it. The remarkable thing, however, is that the centers of gravity of the strips are in a line and will form a median of the triangle. Therefore, the point of application of the resultant will fall somewhere along this median. The same triangle is now divided into strips parallel to a different side (Illustration 4b). The centers of gravity of the strips will form a second median. The point of intersection of the two medians (Illustration 4c) is the center of gravity of the triangle. A third median will intersect the meeting point of the first two medians.

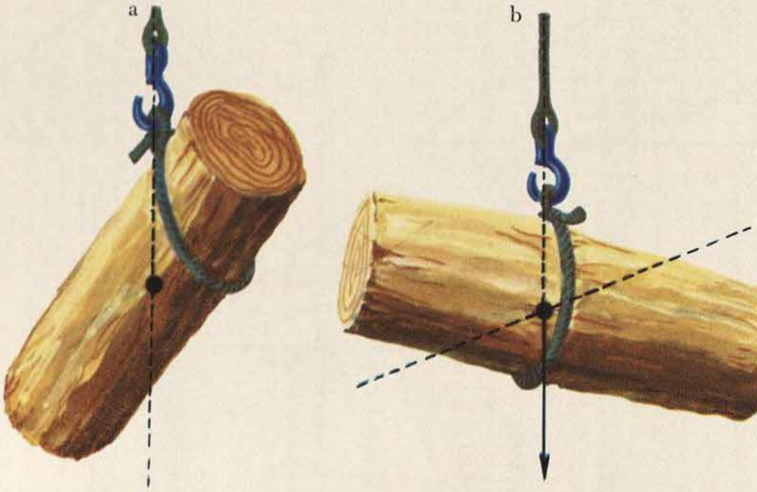


5



THE CENTER OF GRAVITY OF A BODY HAVING COMPLEX FORM—Because of the simplicity of their forms, it has been easy to determine the centers of gravity of the cylindrical iron bar and a triangle. When the body has a more complex form, such as that of an irregular solid, the weight forces of all the elements into which the body might be reduced must be calculated. Several weight forces, for example, must be calculated for the body shown in Illustration 5.

6



AN EXPERIMENTAL METHOD OF FINDING THE CENTER OF GRAVITY—The weight of a body is always applied at its center of gravity; therefore, if the body is suspended at any point (Illustration 6a), it will come to rest in such a way that the center of gravity is directly below the point of suspension. The center of gravity will be located somewhere on a plumb line drawn from the point of suspension. If the same body is suspended at another point (Illustration 6b) and another plumb line is drawn from the point of suspension, the intersection point of the two lines is the center of gravity. When a body is regular in form, its center of gravity is not difficult to determine, but may involve long and laborious calculation. In construction projects, the calculation of centers of gravity is very important and is applied in a vast number of ways.

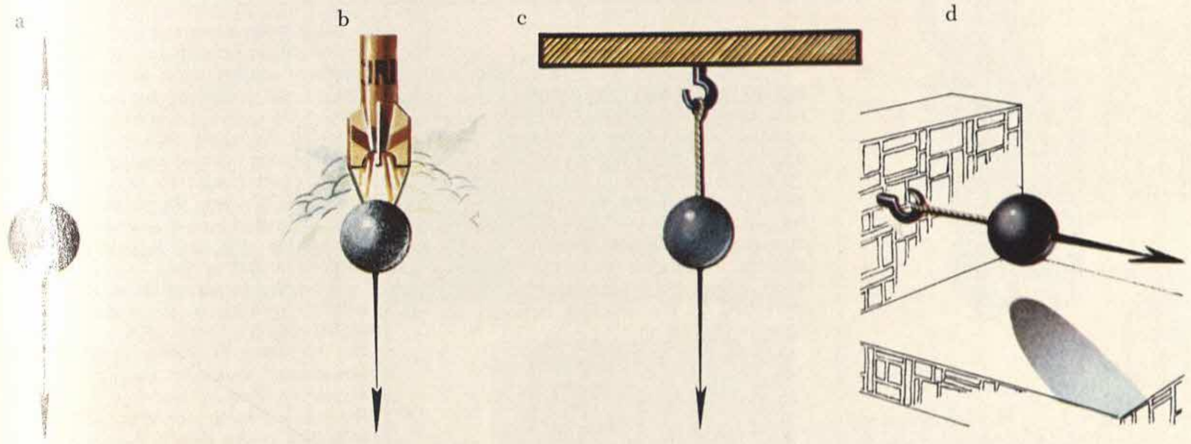
THE EQUILIBRIUM OF FORCES | bodies at rest

It is one thing to recognize and know the composition of the forces that cause a body to move. It is a different thing to arrange these forces so that equilibrium is attained and the body to which they are applied remains at rest. Adjusting these forces is the practical purpose of statics, a branch of the science of mechanics.

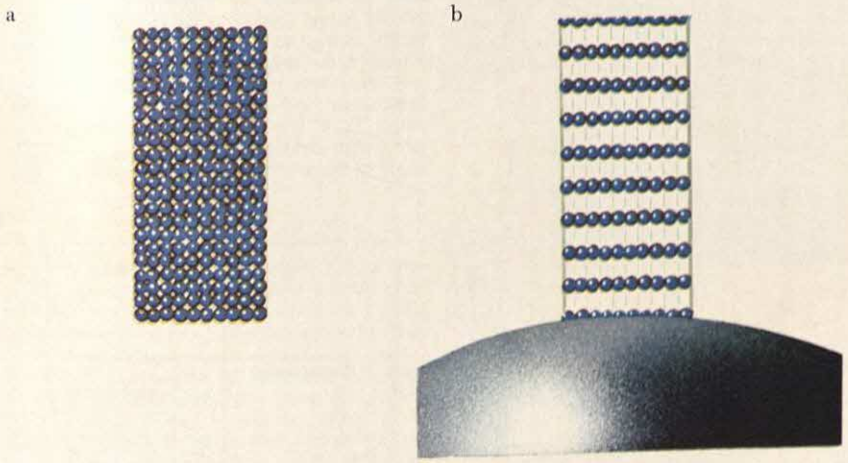
1

EQUILIBRATING A FORCE: THEORY AND PRACTICE—The downward-pointing arrow in Illustration 1a represents an isolated force. It may represent the weight of a body that, if it is not affected by other forces, will move downward. This weight will produce the downward movement unless the force is equilibrated. Equilibrium can be achieved by applying a second force equal to the first, in a direction opposite to the first force. In practice, a way must be found of generating this opposite force (represented by the upward-pointing arrow). One possibility is to use a rocket engine with a thrust equal to the weight of the body (Illustration 1b). The thrust of the engine will be directed upward and the body will remain in perfect equilibrium.

Another way of equilibrating the weight of the same body is to suspend it from a beam by a cord (Illustration 1c). The body will then be held up by the resistance of the beam and the cord. In the previous example, the rocket thrust was an active force; here, the force is passive. A passive force, however, exists only because it is generated by the force it is opposing; the cord exerts a force on the ball it is suspending for as long as the ball remains hanging. If the ball is removed, the force of the cord stops acting. An experiment (Illustration 1d) demonstrates this. A cord is attached to a ball and to a stable support. The ball is then pulled in the direction indicated by the arrow. Once the cord is taut, the ball can no longer be moved in the direction of the arrow; the cord exerts a force on the ball and keeps it at rest. If no force is exerted on the cord, however, the cord will not set the ball in motion in the opposite direction. The force exerted by the cord is a passive force that acts only as a reaction to another force that provokes its action. Such forces are also called resisting forces or binding reactions.

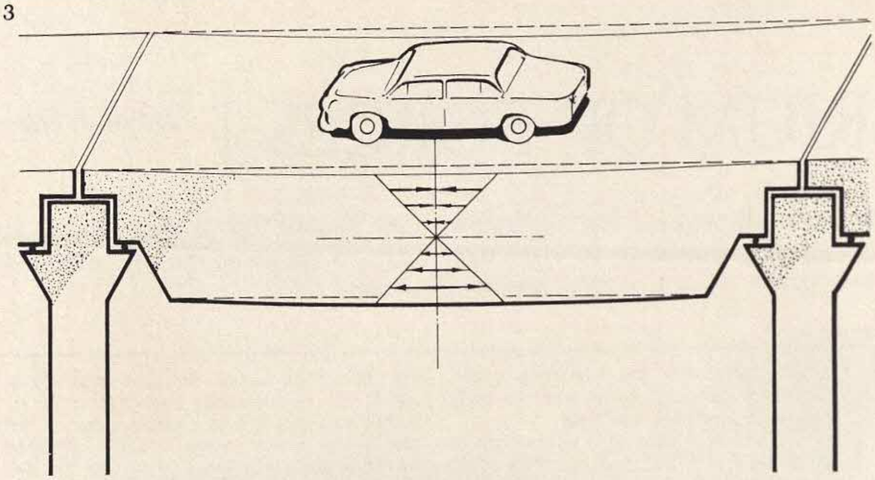


2



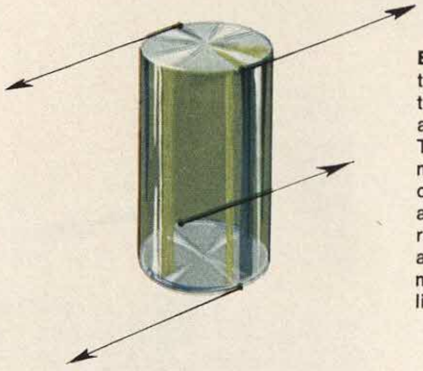
THE ORIGIN OF PASSIVE FORCES—Solid bodies are made up of atoms arranged in fixed positions. If a force acting on these atoms tends to move them apart, the force holding them in their original position reacts. Illustration 2a shows a greatly enlarged view of the atoms in the cord (shown in Illustration 1c) while no force is being exerted. A downward force exerted on the cord tends to separate the atoms composing the matter of the cord (Illustration 2b). The atoms, which are in a state of equilibrium before the downward force is applied, return to their state of equilibrium when the downward force is removed. The downward force, when active, also acts on the atoms of the hook, the support, and all components resisting the downward force.

3



THE PROBLEM OF CONSTRUCTION—This bridge section consists of a beam resting on two piers. When a vehicle crosses the bridge, forces are generated inside each component of the section. The lower part of the horizontal beam is stretched by the weight of the beam itself, plus the load of the vehicle it is bearing. This tension is represented by arrows pointing in opposite directions in the middle of the beam. Resistance to these forces is provided by the passive reaction of the material from which the bridge is made: in this case, iron-reinforced concrete. The engineer who designs such a bridge must calculate the strength of this stretching force, taking into consideration the weight of the beam and the vehicular loads it must bear. He will decide what dimensions to give the beam to ensure that it will not break. The material bonds that connect parts of the structure must also be studied. Bonds can resist forces only up to a certain point, beyond which they break. Calculating accurately the strength required for such bonds as welds, rivets, and bolts is of primary importance to the designer who determines the dimensions of the structure, and is one of the most complex problems of statics.

4



BALANCING PAIRS OF FORCES—Two identical, parallel forces acting in opposite directions are called a couple. When a couple is applied to a body, the body tends to rotate. To keep the body at rest, an identical couple must be introduced to counteract the first couple. Illustration 4 shows a cylinder held at rest by two couples acting in opposite directions. The moment of a couple is defined as the product of the magnitude of each force multiplied by the distance between the two lines of application.

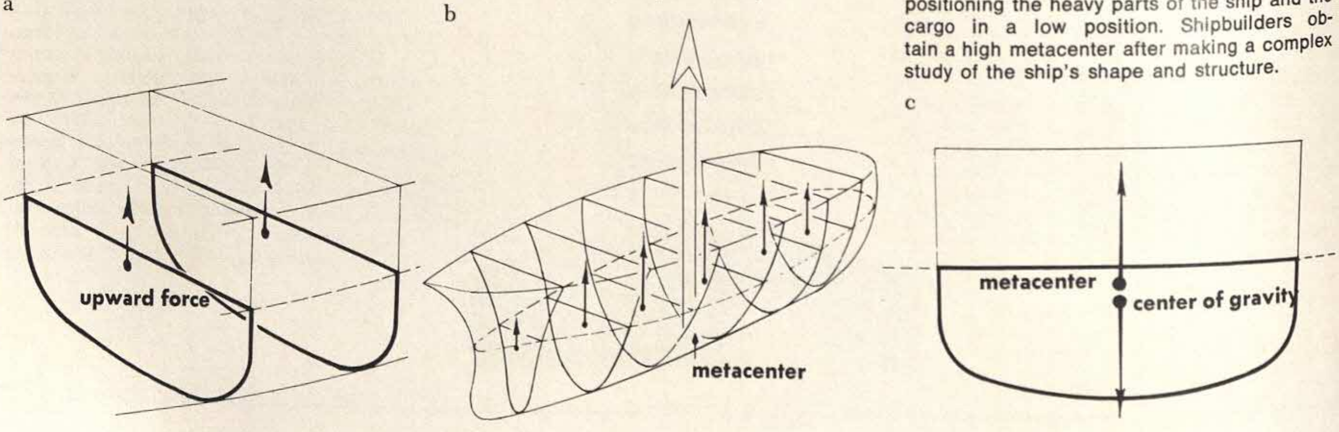
5

THE CENTER OF GRAVITY, METACENTER, AND EQUILIBRIUM OF SHIPS—The hull of a ship immersed in water is subject to Archimedes' principle of buoyancy. This principle involves a force acting upward with an intensity equal to the weight of the volume of water displaced by the ship's hull. At what point of the hull is this force applied? Solving this problem is analogous to that of finding the center of gravity of a body.

Imagine the immersed part of the ship (Illustration 5a) to be divided into sections, or elements of volume, each of which is subject to the upward buoyancy force. The problem of finding the central point of application of this force for the whole ship is equivalent to that of finding the point of application of the forces acting on each element of volume into which the hull is divided. The point of application of all the upward forces is within the immersed part of the ship and is known as the *metacenter* (Illustration 5b).

For a ship to remain in equilibrium in water without overturning, the metacenter must be higher than the center of gravity. In Illustration 5c, the center of gravity and the metacenter are aligned vertically and tend to keep the ship in equilibrium. Both the weight force and the upward force act vertically. Thus, if the ship lists to one side, the couple arising from this pair of forces tends to rotate the ship back to an even keel. The greater the moment of the couple, the more stable the ship. Stability is achieved by making the center of gravity fall as low as possible and the metacenter as high as possible. If the forces are so arranged, the ship is unlikely to overturn.

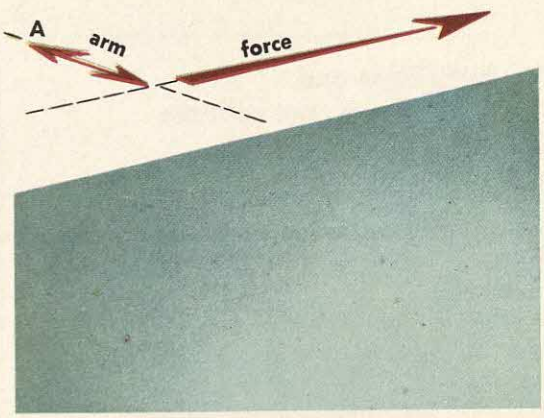
A low center of gravity can be obtained by positioning the heavy parts of the ship and the cargo in a low position. Shipbuilders obtain a high metacenter after making a complex study of the ship's shape and structure.



6

THE MOMENT OF A FORCE IN RELATION TO A POINT—Just as the moment of a couple has been defined, the moment of a single force in relation to a point may also be defined. The moment of a force is the product of the magnitude of the force multiplied by the distance of the point **A** from the line of action of the force (Illustration 6a). When a bolt is tightened with a wrench (Illustration 6b), a force is exerted perpendicular to the arm of the wrench. The moment of the force acting on the bolt is equal to the product of the perpendicular force exerted by the hand multiplied by the length of the handle of the wrench.

a



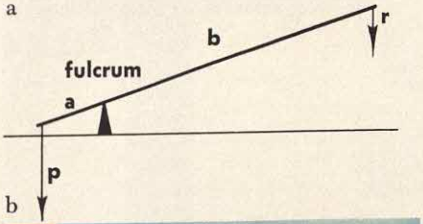
b

7

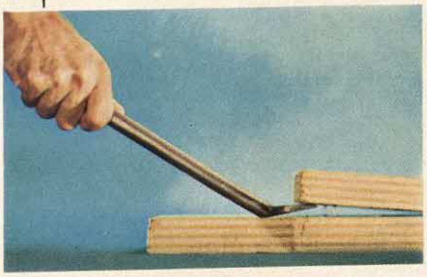
LEVERS, PULLEYS, WINCHES, AND WEDGES—Illustration 7a shows a lever—a bar that rotates around a pivot called a *fulcrum.* If a force is exerted on one end of the lever, the force will have a moment in relation to the fulcrum. If the force **p** is exerted on the end of the arm **a**, the moment will be $M = ap$. If an appropriate force **r** is applied to the opposite end of the arm **b,** the result will be an equilibrant to the moment of the force applied at the other end. This second moment will be $M = br$. If the lever is in equilibrium the sum of the two moments is 0, therefore, $ap - br = 0$, which is the same thing as saying $ap = br$ or that $a/b = r/p$. Thus a lever is in equilibrium when the forces applied to the two ends are inversely proportional to the lengths of the arms. A small force applied to a long arm, therefore, can equilibrate a big force applied to the short arm of the same lever. A crowbar (Illustration 7b) is a lever with arms of greatly different lengths. Suppose one arm is 12 in. long and the other 2 in.; a force of 1 lb on the end of the long arm can balance a force of 6 lbs applied to the short arm: $(12/2) \times 1 = 6$. Levers are described as having a mechanical advantage.

A pulley (Illustration 7c) functions in a way similar to a lever. When a rope is placed around it, the pulley is the equivalent of a lever with ropes at either end and pulling in opposite directions. For the pulley to be in equilibrium, it is necessary for the moments of the forces exerted by the two cords to be equal.

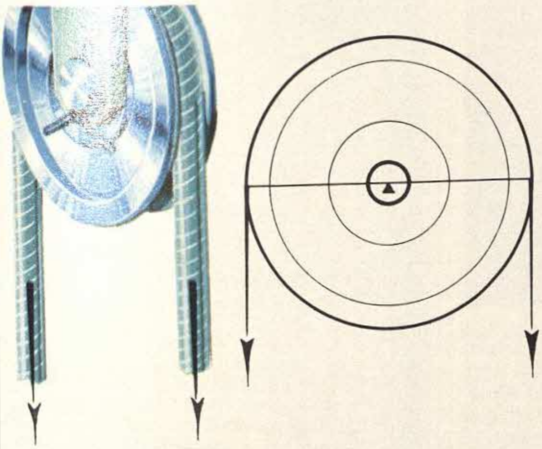
a



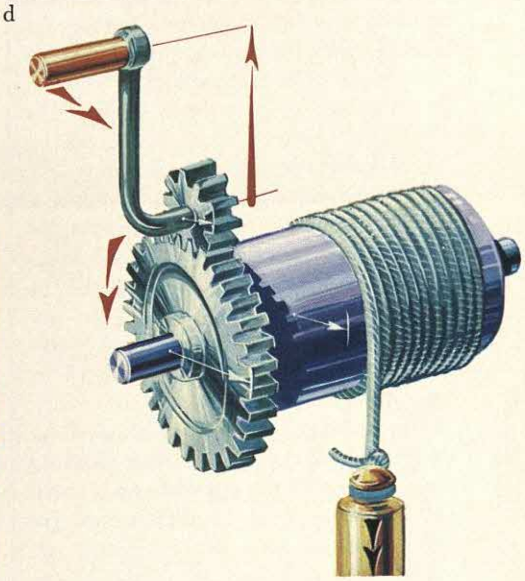
b



c



d

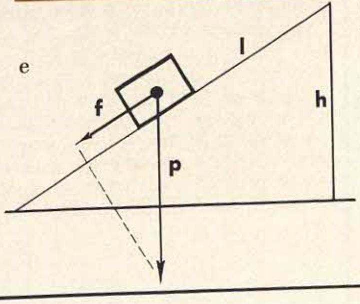


The equilibrium of a winch (Illustration 7d) is attained through a number of levers, each with arms as long as the radii of the gear wheels.

An inclined plane (Illustration 7e) is a machine that has a mechanical advantage. A weight pushed up an inclined plane reacts with a force of a fraction of its own weight. This fraction amounts to $f = p(h/l)$, when f is the force with which the weight is kept in equilibrium, p the weight of the body, h the height of the inclined plane, and l its length. The effort made is thus reduced by the amount of the ratio h/l. The multiplication of forces produced by the machines described does not contradict the basic laws of dynamics.

In all of these machines, movement results if a slightly larger force is exerted than that necessary to maintain equilibrium. Their usefulness lies in the fact that a small force can be applied to move a large force. This usefulness, or mechanical advantage, is accomplished at the expense of having to move the small force through a greater distance than the large force is moved (as, for example, in the case of the lever). The basic laws of dynamics are upheld—mechanical energy, which is the product of the force and the distance it moves, is conserved throughout.

e



STATICS

why bridges and skyscrapers do not collapse

A bold engineering achievement—a great bridge or a towering skyscraper—may show the skillful use of the builder's materials but reveal nothing of the colossal forces imprisoned within the concrete or steel, forces that would destroy the structure if they had not been suspected and predicted. The bridge would fall and the skyscraper collapse if the planners and builders lacked a knowledge of statics—the branch of physics that deals with the conditions under which forces acting on a body create their own equilibrium.

A simple problem in statics is involved in the planning of a small bridge that consists of a metal beam supported by two masonry piers, one at each end. The designer needs to estimate the forces that will be applied to the beam—the weight of the beam itself along with the weight of persons crossing it (represented by arrows pointed downward) and the vertical force applied by the supporting piers (arrows pointed upward) separated by the distance between them.

The beam will be in equilibrium if the resultant moment (force times distance from axis) of all the applied forces about any fixed axis is nil. In other words, the forces applied downward, upward, and outward from the center of the bridge cancel each other out and the body of the bridge remains at rest.

The problem is more complex if the bridge is longer, with three spans supported by four piers. The weight of the side sections exerts pressure on the center section, and the center section exerts pressure on the side spans. These forces are taken into consideration when the bridge is constructed. The removal of one span would cause the collapse of the entire bridge because the change in the forces applied to the remaining spans would destroy the equilibrium of the structure.

A knowledge of the statics of rigid bodies is of critical importance in the design and construction of such varied objects as aircraft, dams, submarines, automobiles, high-heeled shoes, and braces for straightening teeth.

1

FORCES—Statics is concerned with forces originating from different sources—forces also differing in intensity and the direction in which they are applied. Forces are represented on paper by an arrow pointing in the direction in which the force will be applied. The length of the arrow, drawn to a given scale, reflects the magnitude of the force. For example, a cord tied at a fixed point is pulled taut by a hand exerting a force of 5 kg (about 11 lbs) along the straight line from the hand to the point where the cord is fixed (Illustration 1a). The direction of the force is away from this point. If a line 5 mm in length is used to represent a force of 1 kg, an arrow representing a force of 5 kg must be 25 mm long.

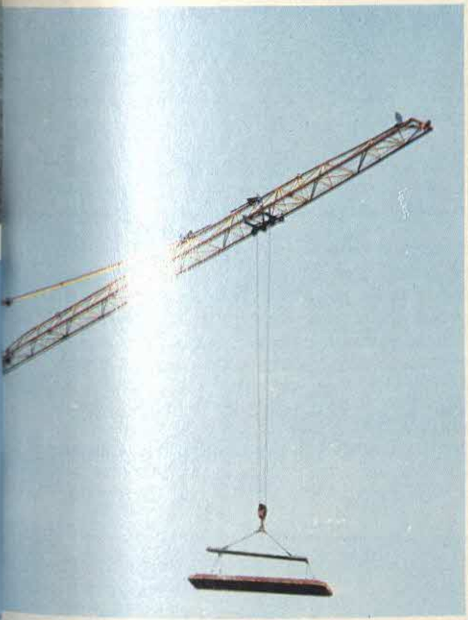
The force of a body resting on a plane is shown by a balance weight of 1 kg exerting a downward force represented by an arrow 1 cm long (Illustration 1b).

The hull of a floating ship (Illustration 1c) is buoyed up, according to Archimedes' law, by a force equal to the weight of the water displaced. In this case, the force, 20,000 tons, might be represented by an arrow 20 mm in length.

Obviously, a heavy loaded ship will ride lower in the water than an empty one. The two ships may be exactly the same size, but the amount of water their hulls displace will vary with the total weight of ship plus contents. Freighters of the type shown carry ore, coal and grain on the Great Lakes.

The wind blowing at a certain velocity against the building in Illustration 1d exerts a force of 1 kg/m^2 of surface. The facade has an area of 6,000 m^2 and is therefore subjected to a force of 6 metric tons, which could be represented by an arrow 6 mm long.

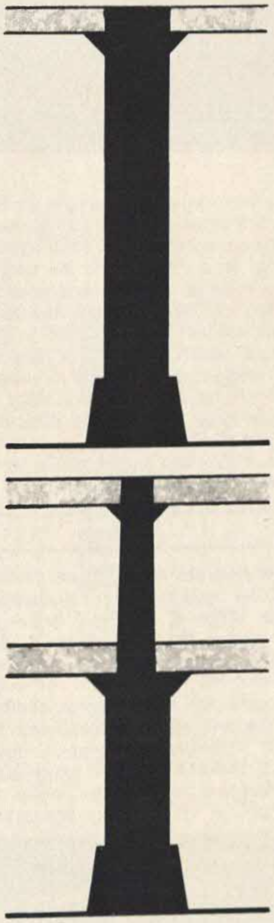
e



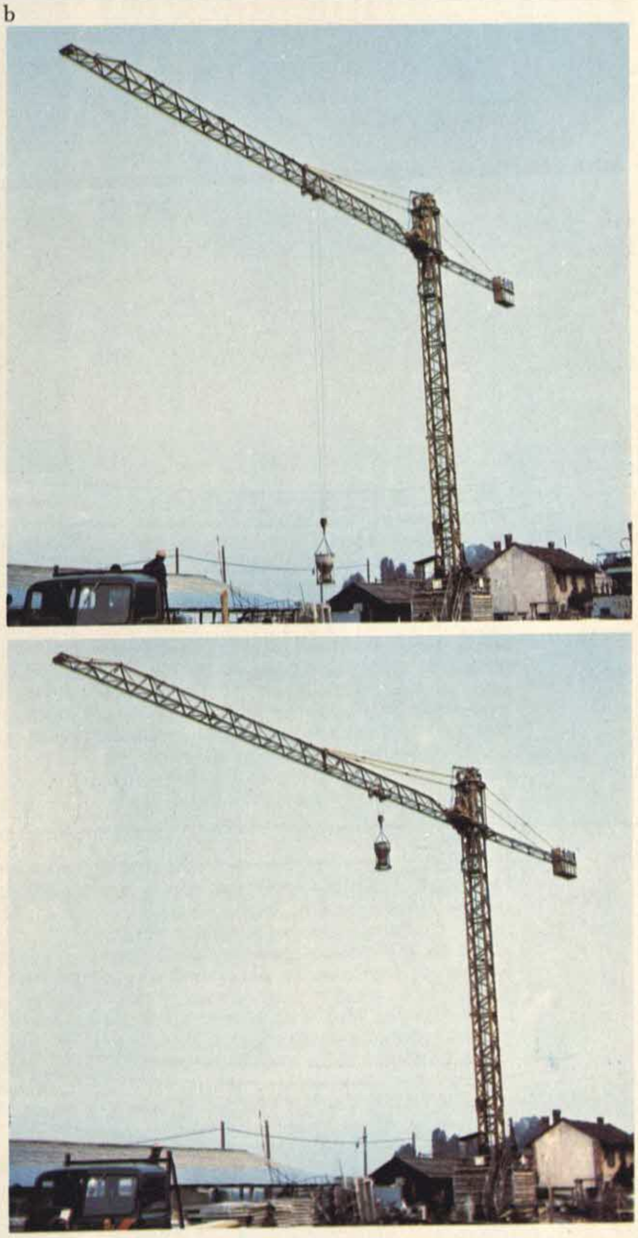
The upper arm of the crane (Illustration 1e) is subjected to a force that could be represented by arrows. The amount of the force can be determined by calculation that includes the weight of the arm, the length of the arm, and the weight of the load being lifted. If the top supporting strut were cut, the arm would break under its own weight or the weight of the load. The knowledge of the forces to which such equipment is subjected at every point is necessary to the efficient operation of that equipment.

2

a



b

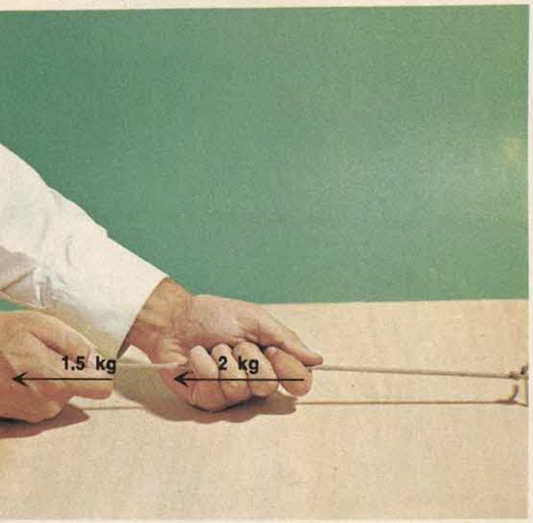


THE POINT OF APPLICATION OF A FORCE— In some cases it is necessary to know the exact point of application of a force. The bridge pier in Illustration 2a has shelves at a certain height supporting weights on each side. The force with which the pier presses on the Earth is the sum of the weight of the pier plus the weights in the shelves. If the pier is designed to accommodate additional shelves, the part of the pier that extends from the additional shelves to the original shelves may be decreased in size. It is important to know exactly what weight will be exerted by the shelves at any given height in order to determine the dimensions of the pier.

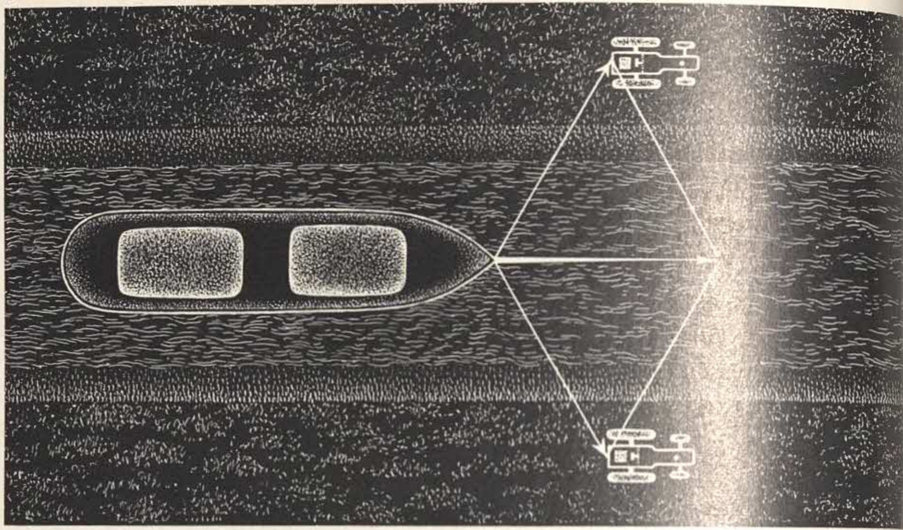
While the crane (Illustration 2b) is lifting a weight suspended from its arm by a cable, the weight moves upward or downward depending on whether the cable is being shortened or lengthened. However, the stress to which the top supporting strut of the arm is subjected remains constant. In a majority of cases in statics, the point of application does not matter. The magnitudes of forces can be shown by arrows along the lines of their application. In statics, it is often convenient to move forces along their lines of application in order to solve problems more easily.

3

a



b



THE SUM OF FORCES—It is highly important in statics to be able to calculate the sum of two different forces. For example, in Illustration 3a, a cord tied to a hook is drawn taut. The first hand is exerting a force of 2 kg (about 4.4 lbs), represented by an arrow 2 cm long. The second hand is pulling on the cord with a force of 1.5 kg (about 3.3 lbs), represented by an arrow 1.5 cm long. The total force exerted on the hook, 3.5 kg, may be determined by placing the arrows that represent the two forces head to tail. The combined length of the arrows, 3.5 cm, represents the sum of the forces acting on the hook. This example illustrates the sum of concurrent parallel forces.

In a slightly more complex example of the sum of two forces (Illustration 3b), two tractors are pulling a barge along a canal, each tractor exerting a force on a rope tied to the barge. The two ropes meet at an angle and each is exerting a force of 300 kg (about 660 lbs). The two forces are not being exerted in the exact direction in which the barge is moving, but each at an angle of 60° from that direction. The total pulling force acting on the barge is made up of one force acting in the direction of the movement of the barge and a second force acting in a direction perpendicular to it. The forces perpendicular to the direction in which the barge is moving are ineffective as far as its movement is concerned, but the forces acting with the direction of the movement help to pull the barge forward. It can be said that these forces have been projected toward the direction of movement (because the arrows representing them are obtained from the original forces by a geometrical projection), or that the two forces have been reduced and only the components acting parallel to the direction of movement are being considered. Therefore, the force exerted on the barge is equal to the sum of the two components, each coming from the pulling force of one of the tractors.

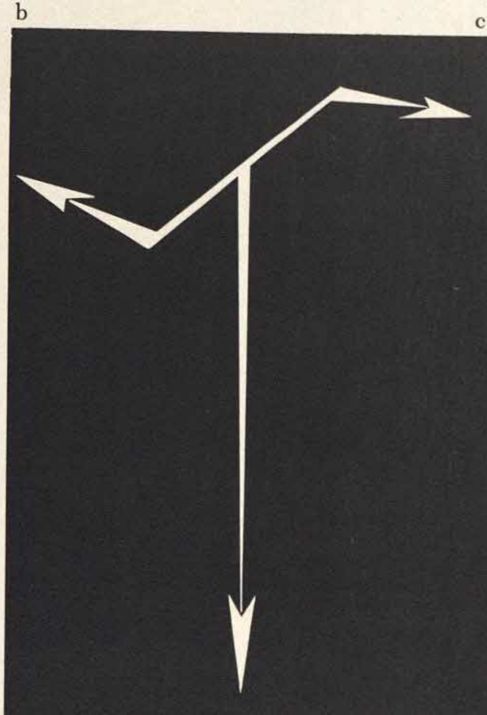
4

PAIRS OF FORCES—When two forces are equally strong, parallel, and exerted in opposite directions, their sum cannot be obtained by the method shown in Illustrations 3a and 3b. They must be considered as two distinct forces together exerting what is called a pair of forces. When a corkscrew is screwed into the cork of a bottle (Illustration 4a), the thumb and index finger exert a pair of forces, causing the corkscrew to rotate. While a single force may cause movement of a body, a pair of forces that are not balanced causes the rotation of the body to which it is applied.

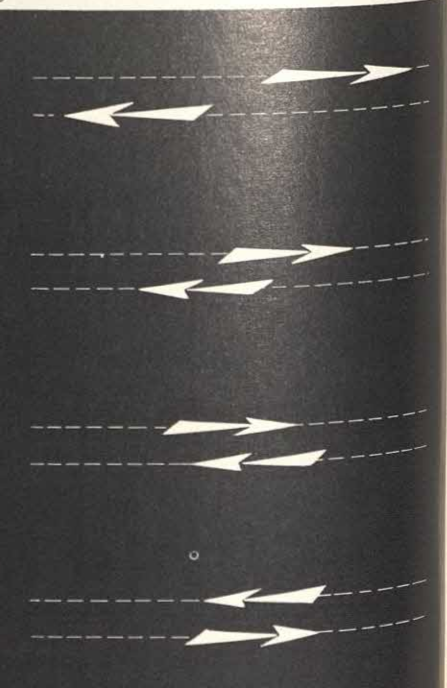
The effective action of a pair of forces is described as torque, the name given to the product of the intensity of each of the two forces multiplied by the distance between them. The torque of the pair of forces needed to turn a corkscrew may be calculated by multiplying the force exerted by each finger by the distance between them. (In expressing torque, the distance between the forces is given first so as not to confuse the torque with a measurement of energy.) Illustration 4b shows how the forces are arranged when a pair of forces is operating a corkscrew. The distance between two parallel forces is known as the arm. The examples in Illustration 4c show how it is possible by beginning with a single pair of forces to obtain an infinite number of pairs, all with the same torque. It is sufficient to allow each force to run along the line of application.

a

b



c



OSCILLATING MOTION | vibrations and resonance

1

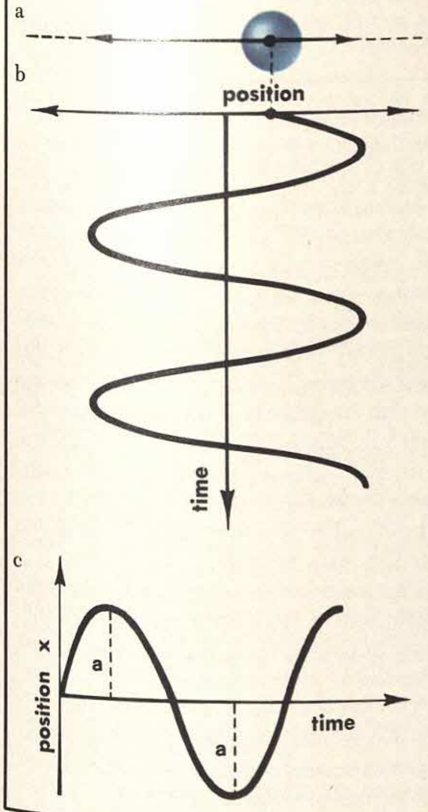
SIMPLE HARMONIC MOTION—In order to present a good kinematic definition of any motion, it is helpful to describe the motion of a point. A point that carries out a harmonic oscillating motion will periodically move forward and backward along a segment of a straight line (Illustration 1a). This means that if the point at a certain instant of time occupies a certain position on the segment and has arrived at this position, say from the left end of the segment, it will return to occupy this same position, again coming from the left, in subsequent instants. These instants can be determined from the initial one by adding a fixed interval of time known as the period. The law that governs the movement of the point along the segment is schematically described in the following illustrations.

Suppose that the point carrying out this motion can trace a line on a strip of paper that is moving at a right angle to its direction of oscillation (Illustration 1b). If the paper strip is moving at a uniform speed, then the line traced by the oscillating point will be a sinusoid. It can be deduced from this that in simple harmonic motion the speed is zero at the two extremes of the segment and attains its maximum value at the center.

The oscillating motion can be described mathematically as follows: The point moves along a segment of length **2a,** which lies on the x axis with its center at the origin of the coordinate system (Illustration 1c). If the origin of the time scale coincides with one of the instants at which the point finds itself at the center of the segment, then the motion of the point is described by the formula

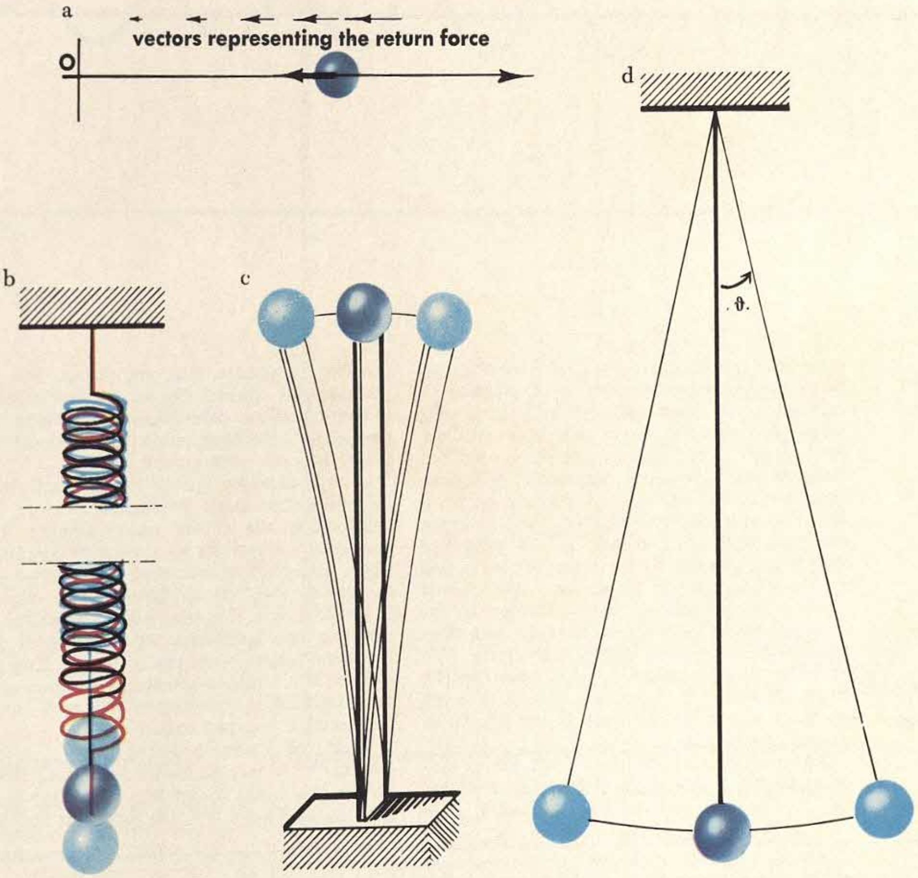
$$x = a \cos \frac{2\pi}{T} t.$$

In this formula, T is the period of the oscillation, and t represents the time that has elapsed from the initial instant when $t = 0$.



A mechanical problem of fundamental importance for the whole of physics, engineering, and technology is that represented by vibrations. The mechanical problem can be formulated in the following manner: given a body that has a certain mass and is subject to certain forces, can its motion be a periodic one, like that of a pendulum? If the answer is in the affirmative, with what frequency and maximum amplitude will this motion occur? Or, if an oscillating motion is damped,

2



DYNAMIC CAUSES OF HARMONIC MOTION—In Illustration 1, a kinematic description of harmonic motion was given. The next problem is to see what physical circumstances cause harmonic motion.

A body (Illustration 2a) that is represented by a material point can move along a straight line and is subject to a return force that always acts in the direction of the point **O.** The intensity of this force is proportional to the displacement from **O;** it is, therefore, zero at **O** itself. A force that has these characteristics is of the elastic type and gives rise to a motion that occurs with the maximum acceleration at the two ends of the travel where the force attains its maximum intensity, while both the force and the acceleration are zero at **O** itself.

An example of elastic forces is provided by a spiral spring. Illustration 2b shows a body attached to such a spring. If the body is allowed to remain at rest, it assumes a position in which the force of the spring and the force due to the weight of the body are in a state of equilibrium. If the body is raised, thereby compressing the spring, the body will be pushed downward by a force that, within certain limits, is proportional to the shortening that has been produced in the spring. Subsequently, the body will pass the equilibrium point with zero acceleration, but with a velocity that is directed downward. As it goes beyond the equilibrium point, the body lengthens the spring, which will now exert an increasingly intense return force in an upward direction. The motion will thus continue in this manner.

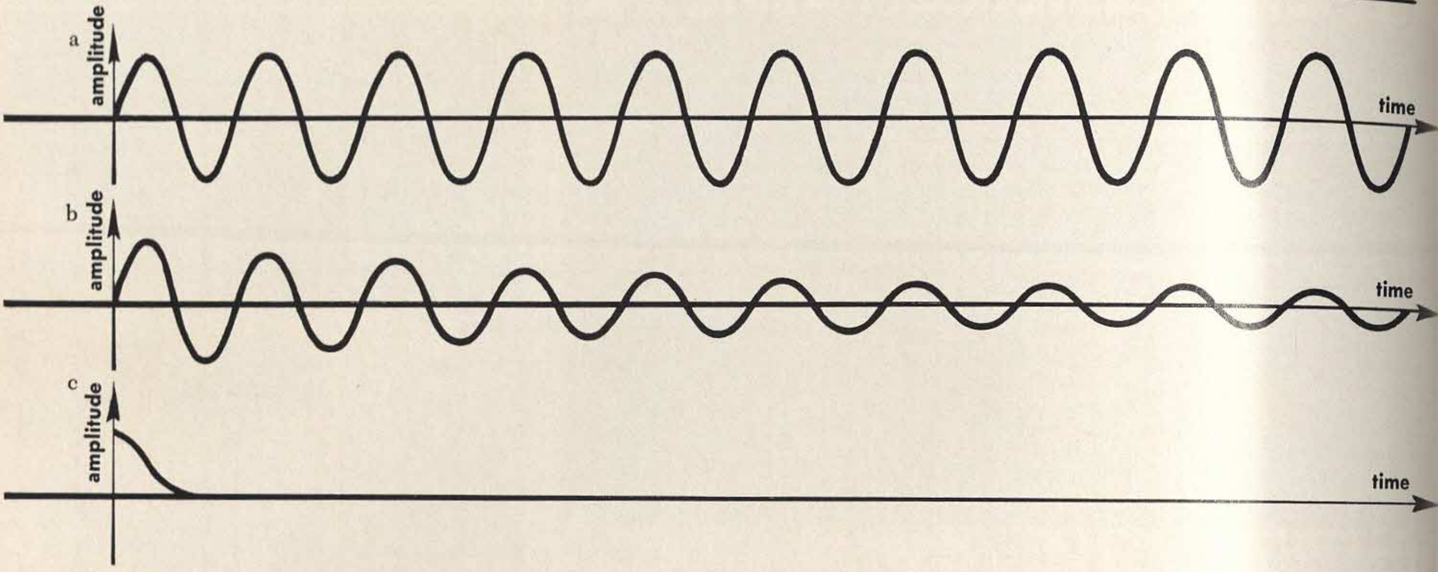
Another example of an elastic force occurs in a vibrating lamina that carries a weight at its free end (Illustration 2c). The force that tends to pull the suspended mass back to the rest position is proportional to the amount of displacement of the mass from this rest position.

The mass of a pendulum is subject to the return force of the weight; this force, as shown in Illustration 2d, varies according to the angle that the pendulum forms with the vertical. If P is the weight of the pendulum and θ is the angle between the pendulum and the vertical, this force will be given by $F = P \sin \theta$.

When the angle between the pendulum and the vertical is small, the force may be considered to be proportional to the horizontal distance of the mass from its vertical rest position.

In this case, the force may again be said to be an elastic one and the pendulum carries out harmonic oscillations.

3



DAMPED OSCILLATIONS—The harmonic oscillating motion presented in Illustration 1 represents an ideal case. It will occur only when the oscillations can take place without any damping; in practice, this means without friction. Such conditions never exist in macroscopic systems because friction is always present. For example, in the case of the spring a damping action is caused by the imperfect elasticity of the metal; the pendulum is subject to the resistance of the air and the internal friction in the supporting wire which is deformed. These resistances produce a diminution of the amplitude of the oscillations. The formula in Illustration 1, $x = a \cos (2\pi t/T)$, must be modified to include a term that will produce a gradual decrease in the amplitude of the oscillations.

By applying to each of the oscillating systems shown in Illustration 2 an artificial damping action that can be varied at will, it is possible to ensure that the motion will be progressively braked. By varying the amount of braking effect, different laws governing the damping of oscillating motion are obtained and these laws are shown here.

In zero damping the motion is represented by a perfect sinusoid (Illustration 3a).

Illustration 3b shows weak damping. The motion is represented by a sinusoid contained within an envelope consisting of two exponential curves that are symmetrical with respect to the time axis; the exponential curves represent the law governing the damping of the maximum amplitude of the oscillation. The oscillation will become completely damped after a long period of time and its maximum amplitude will be reduced to zero.

When the damping becomes very strong, the oscillation will no longer take place (Illustration 3c). Those forces that attract the body toward the origin are now opposed by very strong forces and the body moves toward the equilibrium position as if it were immersed in a viscous medium. In this case the motion is said to be aperiodic. If the damping is of sufficient strength that the motion of the material point would be at the limit between an aperiodic motion and an oscillating one, the applicable law of motion would be as illustrated.

In this case the damping is said to be critical. Critical damping is characterized by the fact that it damps the motion in the shortest possible time. In fact, if the motion is only weakly damped, the point will carry out a large number of oscillations before it comes to a stop. If, however, the motion is aperiodic, the damping is so strong that a long time will elapse before the point returns to its equilibrium position.

The determination of the critical damping value is of very great importance in practical problems.

what laws govern the extinction of the motion?

It is not difficult to solve this problem in the case of a pendulum or a heavy body attached to a spring, situations that occur frequently in mechanical systems that are of practical interest; however, problems that are more complex frequently occur. For example, a diatomic molecule can be represented schematically as two masses (the two atoms) united by a spring. In this situation, the molecule begins to vibrate when it is struck; that is, the two atoms of which it is made up will alternately move away from and toward each other. The damping of this oscillation, which takes place according to rather complex laws, can be rendered visible outside the molecule by means of the emission of infrared radiation, and the measurement of this radiation will result in data regarding the constitution of the molecule itself.

Civil and mechanical engineering are two other fields in which the study of oscillation is of fundamental importance. A construction in reinforced concrete, such as a bridge, is considered to be a typical static structure—a system that does not move. In practice, however, there is no such thing as an undeformable mechanical system. Every system can become deformed under the action of a stress caused by a sufficiently great force. When these forces cease to act, the system returns to its rest position by carrying out oscillations about this rest position; when the deforming force is itself an oscillating one, the amplitude of the oscillations can increase to the extent that it causes the failure of the structure. This is known as the phenomenon of resonance. It is because of this phenomenon that a convoy consisting of identical heavy vehicles crosses a bridge in such a way that the individual vehicles proceed at different speeds and different distances from each other. If this is not done, the convoy can induce oscillations and, in some cases, oscillate the bridge structure in resonance.

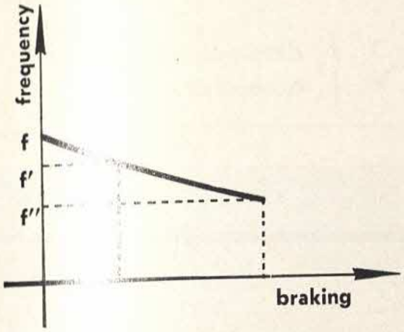
An example of the most complex case is that of a machine with parts that rotate at high speeds. The vibrations to which the rotating parts are subject in some constructions are capable of breaking the machine itself. The designer must, therefore, know the frequencies at which the weakest parts may be caused to vibrate so that resonance phenomena can be avoided. For example, in the case of a turbine it is essential to know the frequencies at which the various parts of a rotating shaft between any two adjacent bearings may vibrate.

The calculation of the vibrations in these cases is complex, requiring complicated formulas that can only be solved with the help of computers.

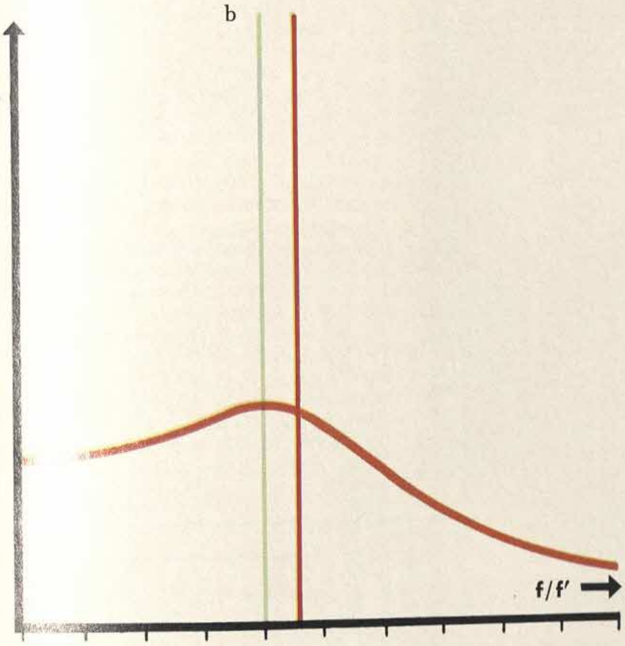
The illustrations and their accompanying discussions show the more important types of oscillating motions and examine the nature of resonance.

4

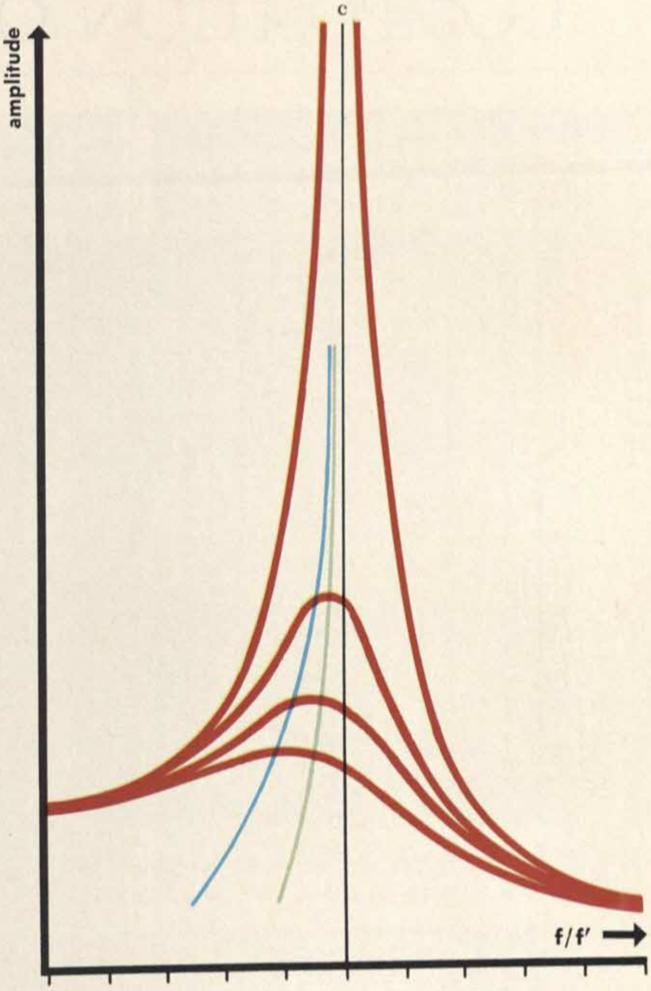
a



b



c



RESONANCE—The foregoing discussions covered free vibrations—that is, vibrations that were allowed to proceed spontaneously after they had been caused. In many cases the oscillations are caused by an external agent—by a mechanical system linked with the one that is oscillating. For example, the vibrations of the body of an automobile caused by the vibrations of the engine are forced vibrations. Forcing a body to vibrate entails pushing it periodically.

Such periodic pushing can give different results. What occurs will depend on the damping of the system being subjected to the forced vibration, and also on the relationship between the frequency of the vibrations of the inducing system and the frequency that characterizes the system in which the vibrations are induced.

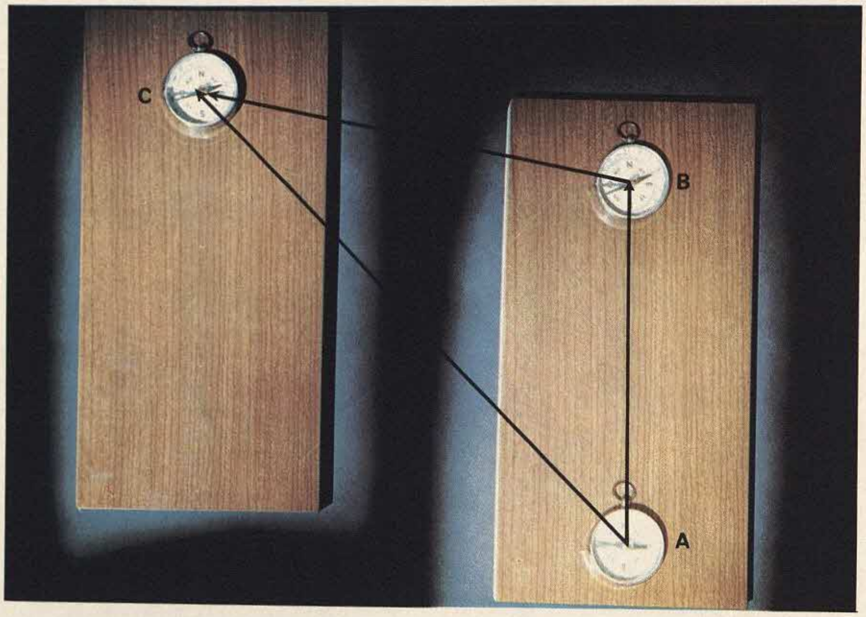
The oscillations depicted in Illustration 4a are free. The characteristic oscillation frequency of the system is **f**, which is marked on the ordinate. However, if a braking action is introduced, both the frequency and the amplitude will diminish, and the frequency now becomes **f′**. If the braking action is further increased, the corresponding frequency **f″** will be even smaller: the greater the braking action, the smaller the frequency. As a parallel, an increase in atmospheric pressure causes a pendulum clock to slow down because the oscillations occur with a smaller frequency and the period increases: the atmospheric pressure brakes the pendulum motion.

Illustration 4b represents an oscillating system with a certain braking action subjected to forced vibration, the period of which can be varied at will. The ratio of the variable frequency **f** of the force that induces the oscillations and the frequency **f′** of the oscillations themselves is plotted. The vertical red line indicates the frequency with which the system would vibrate if it were not damped. However, because the system is damped, its frequency is actually less, as indicated by the vertical green line. If oscillations of varying frequencies are induced, as the frequency of the inducing force increases, the amplitude of the induced vibrations increases until it reaches a maximum at a value slightly less than the resonance frequency in the case of damping. If the frequency is increased still further, the amplitude diminishes again.

Illustration 4c, which summarizes the resonance factor, represents some possible vibrations in a system in which the braking action can be progressively diminished. The black line represents the value of the resonance frequency of the system in the complete absence of damping; the inclined green line, on the other hand, unites all the points of maximum frequency of the oscillations of the system for the various degrees of damping. The blue line indicates the position of the maximum resonance frequencies as the inducing force varies. Examination of one of the curves reveals that the closer the inducing frequency approaches the resonance frequency, the greater the amplitude of the induced vibration. When the damping is zero, the amplitude at the resonance frequency is infinite. In fact, in this case the inducing force continues to supply energy to the vibrating system, pushing it in exactly the same direction in which it is pushed by the spontaneous vibration. The suspension of an automobile, when deprived of its shock absorbers, is a system that is almost wholly devoid of damping. If a car with such a rigid suspension were to encounter a periodic obstacle—for example, the irregularities in the surface of a stone-paved road where all the paving stones are of the same dimensions—the car would begin to resonate and the amplitude of its vibrations could become enormous. In practice, the vehicle would be unable to hold the road. Because of the great amplitude that the vibrations can assume as a result of periodic obstacles that constitute a vibration-inducing force, the suspension of a car capable of high speeds must be provided with very carefully designed damping. The irregular forces that the surface defects of a road impress on the wheels of a car have a white spectrum; that is, they contain all the frequencies of vibration, including the ones best suited to being amplified by the oscillating system of the suspension. The shock absorbers thus must always be designed to provide critical damping. The design of the shock absorber system of a vehicle can be easily and rapidly carried out with the help of analog computers that are capable of perfectly simulating the various conditions of oscillation of the vehicle and its suspension system when they are subjected to stresses induced by the road.

COMPOSITION OF MOTION | displacement, velocity, acceleration

1

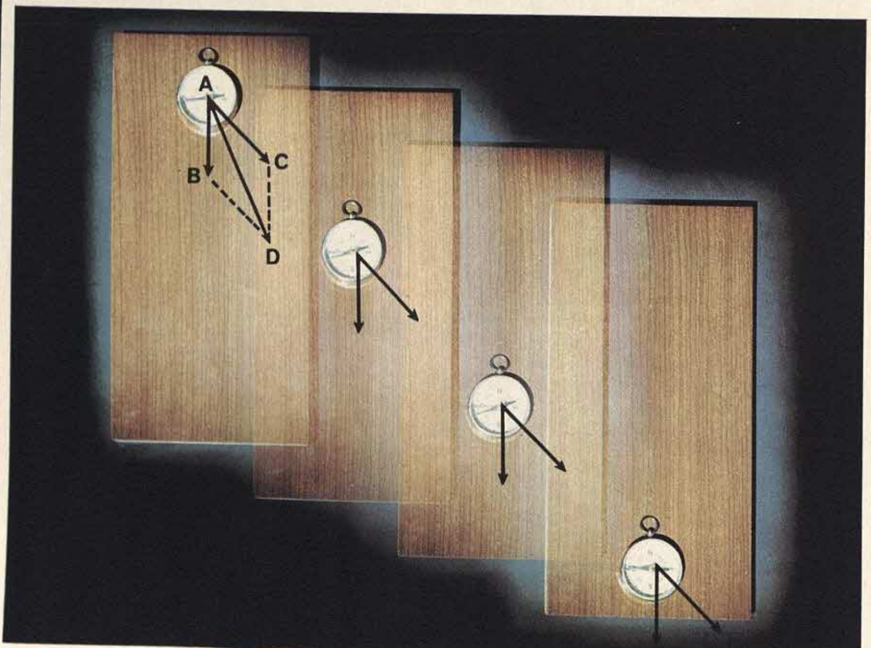


THE COMPOSITION OF DISPLACEMENTS—Before the problem of the composition of accelerations can be addressed, consideration must be given to the problem of the composition of displacements and velocities. This illustration shows a compass resting on a board, which itself is resting on a plane. The compass was displaced from point **A** to point **B** on the board. Subsequently (things would not change if the displacements had occurred simultaneously), the board was displaced toward the left. The result of these displacements was to bring the compass into the position marked **C.**

The geometric law of the composition of displacements is rather simple. Suffice to say that they can be summed as vectors. The rule of the parallelogram, therefore, applies: the sum of the displacements is given by the third side of the triangle whose other two sides represent the two displacement vectors.

The two sets of coordinate axes are not represented in this illustration. One axis is fixed with respect to the table, which is called reference system O, or the absolute reference system. The other axis is fixed with respect to the board, which is referred to as reference system o, or the relative reference system. The displacement along **AB** is due to the motion of the body with respect to reference system o and is the relative displacement. The one along **BC,** attributable to the displacement of reference system o with respect to reference system O, is referred to as the drag. The displacement **AC** with respect to reference system O is called the absolute displacement, because it is given by the sum of the two component displacements. These definitions are carried over into Illustration 2, which further explains the composition of motion.

2

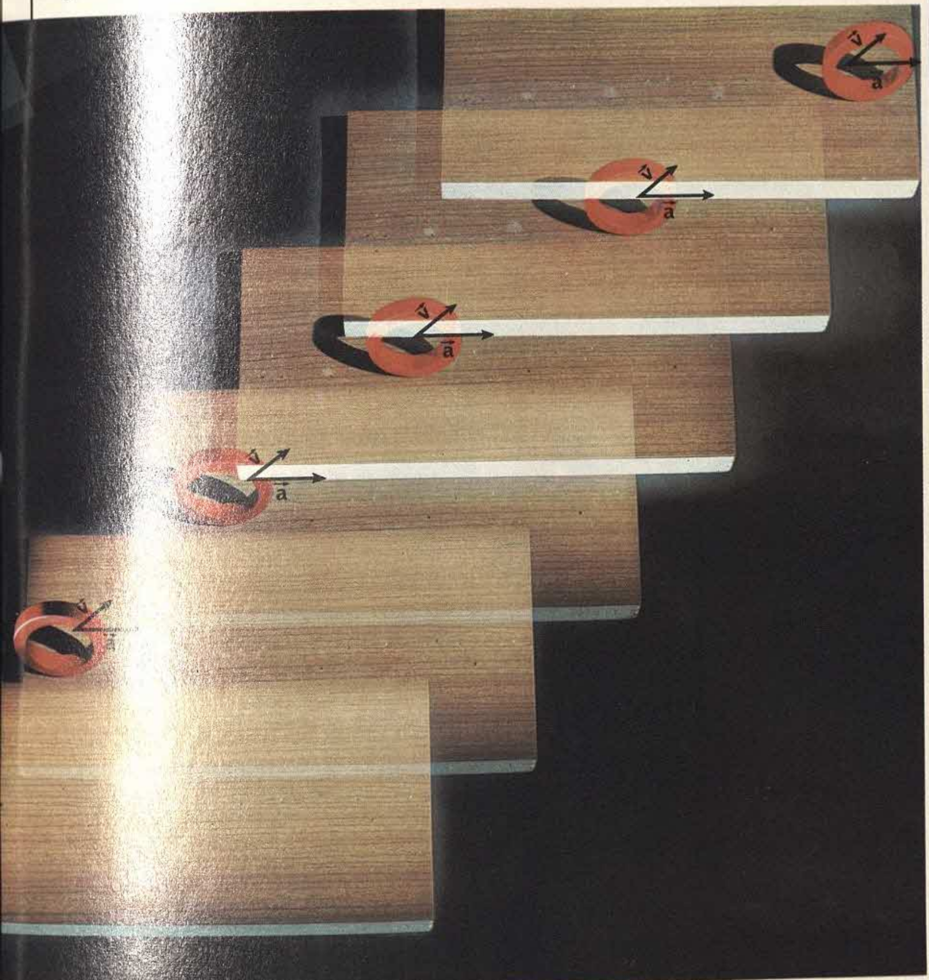


THE VECTOR SUM COMPOSITION OF VELOCITIES—As in the case of the displacements, velocities are summed by adding the vectors that represent them. This illustration shows a compass fixed to a board that is moving with respect to the table underneath it. The compass is then moved with respect to the board. Because the motion is considered over a very brief length of time, its velocity is said to be constant.

The same principle applies to the board, which is sliding with respect to the table. Considered for an instant, its velocity is said to be constant. Both movements are represented by means of vectors. The resultant velocity is the sum of the two movements.

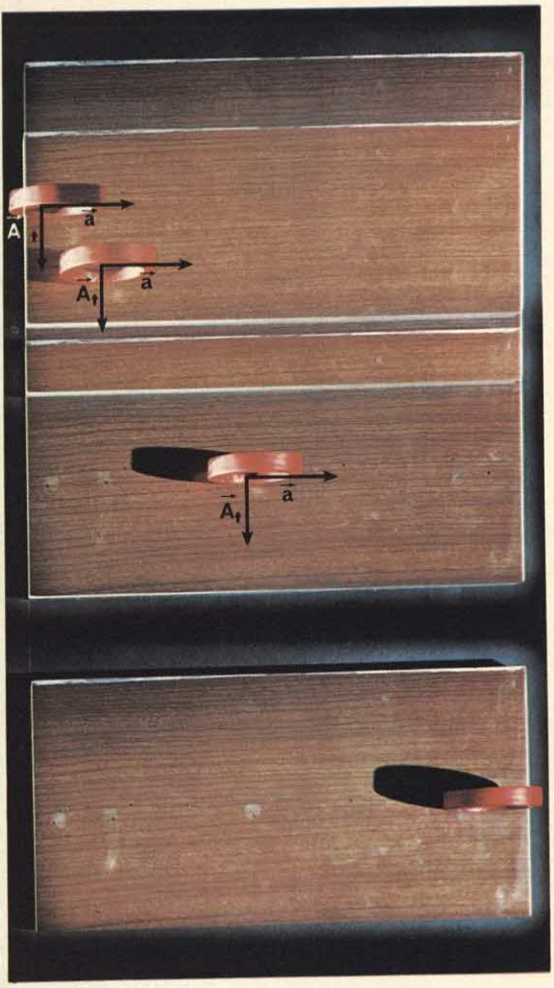
The nomenclature is identical with that used in Illustration 1; however, the illustrations are different. In Illustration 1, the length of the arrows represents the length of a displacement; here, it represents the value of a velocity. In Illustration 1, when the movement was completed, the body was found to be exactly at the point of the arrow that represents the displacement; here, the arrow merely indicates the instantaneous direction of the movement. If the velocity were not uniform, the body could be found at a point other than that marked by the tip of the arrow. In this illustration, **A** is the point of departure; **AB** is the relative velocity (the velocity of the body with respect to the reference system that is fixed with respect to its immediate support). **AC** is the velocity of the drag (that is, the velocity with which reference system o—fixed with respect to the board—moves with respect to reference system O, which is fixed). **AD** is the sum of the other two velocity vectors.

3



A PARTICULAR CASE OF THE COMPOSITION OF ACCELERATIONS—Accelerations are composed in the same way as velocities and displacements, only in particular cases. The following problem is to be solved here: a body moves with a certain acceleration with respect to a Cartesian reference system, which, in turn, moves with respect to another reference system. What acceleration will the body have with respect to the other reference system? Examination of this problem begins by considering a particular case in which the body moves with acceleration $\vec{a}$ with respect to reference system o consisting (for example) of a system of Cartesian reference axes fixed with respect to the board on which the motion occurs. This reference system (in practice, the board) moves with a rectilinear and uniform motion with respect to the table on which the motion occurs. What acceleration will be observed by an observer who remains fixed with respect to the plane of the table? In this example, the acceleration is the same for both cases. If $\vec{A}$ is the acceleration with respect to the plane of the table, or the absolute acceleration) then $\vec{A} = \vec{a}$. This result is rather intuitive when it is considered that the acceleration with respect to the plane of the table is no different from that on the board because it is seen through a change of uniform translation velocity that cannot contribute to that of the moving body. This case applies to many real phenomena in which a uniformly accelerated rectilinear motion is combined with a movement that occurs at a uniform velocity. For example, any motion on the surface of the Earth becomes compounded with the velocity of the Earth. The latter can be considered to be constant only as a first approximation; in reality, accelerations due to motions of rotation and revolution occur, but they are normally neglected because they are small.

4



A MORE GENERAL CASE—A more general case of the composition of accelerations is illustrated here. A body moves with an accelerated motion with respect to the moving reference system (the board on which it is supported). This board is accelerated with respect to the plane of the table that supports it. However, the acceleration that is impressed on it causes it to carry out a translatory motion with respect to the plane of the table; that is, each point of the board describes a straight line on the plane of the table.

In this case, the composition of the accelerations is effected in accordance with the simple law of the parallelogram. In fact, it can be determined by vector calculation that $\vec{A} = \vec{a} + \vec{A}_t$, (where $\vec{A}_t$ is the acceleration of the drag to which the board is subject with respect to the plane of the table). The absolute acceleration is thus the sum of the relative acceleration and the drag acceleration.

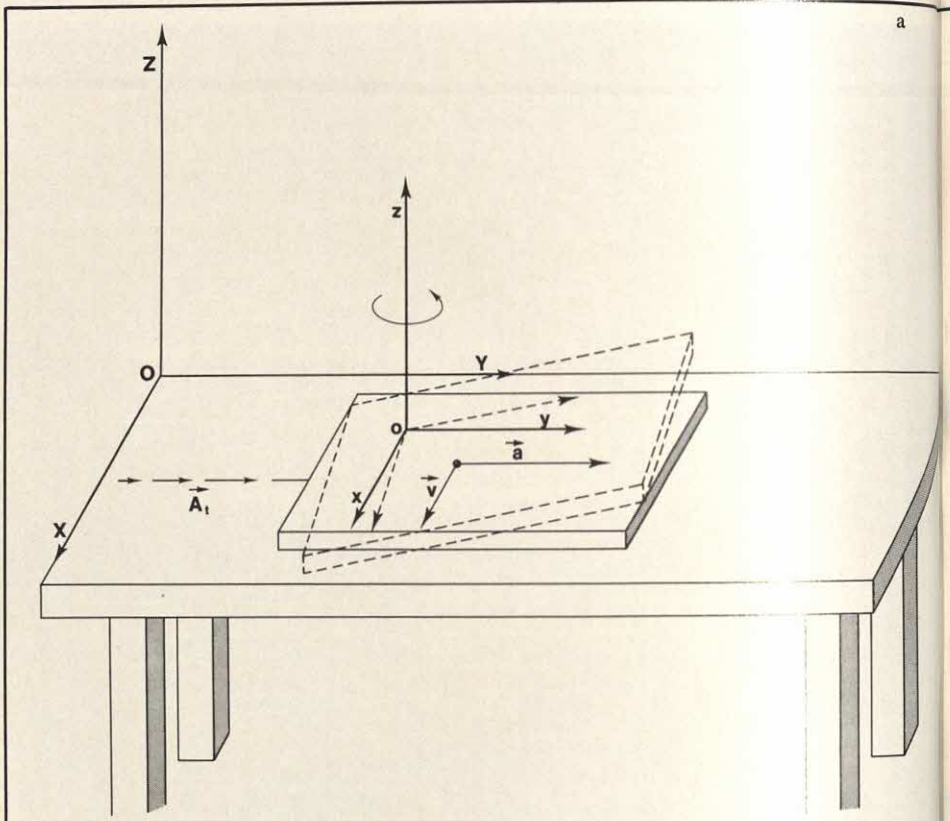
The motion of a body can be described by equations constructed in such a way as to refer the motion of the body to a system of Cartesian coordinate axes. The coordinate system chosen is usually that which is most convenient, in which the equations of the motion are as simple as possible. The three reference axes can be chosen so that two of them describe the plane on which the body is resting, referred to as "reference system o." If this plane (which may be considered to be the loading floor of a trolley) is itself in motion, the body under observation can also be considered to be in motion with respect to another coordinate system—the Earth as a whole. This is a more general reference system, referred to as "reference system O."

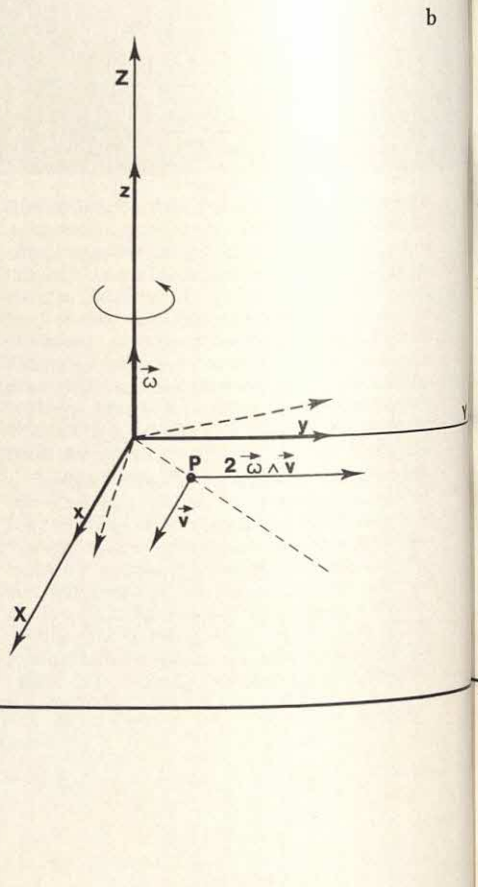
From these two understandings, the velocity and acceleration equations can be obtained. For example, the motion of the Earth is conveniently described in relation to the sun, considering that the center of the sun is reference system o. This makes it possible to find the laws that govern the motion of the Earth around the sun. Once these laws are found, the motion of the Earth can be described with three coordinate axes—fixed with respect to the system of fixed stars. This satisfies the motion of the Earth as regards the reference system o (two axes through the center of the sun). However, the sun itself is in motion among the stars. The larger coordinate system within which the Earth moves is, therefore, O—the movement of the Earth about two axes through the sun, which itself is in movement within the galaxy.

In respect to reference system o, the Earth moves in an ellipse about the sun. In respect to reference system O, the Earth moves in an oblique cylindrical helix wound about an elliptical cylinder.

The difference between o and O as reference systems is not a difference of position; rather, it is a difference derived from the fact that in one case (o) the central reference is considered fixed, and in the other (O) the central reference is recognized as being in movement.

5



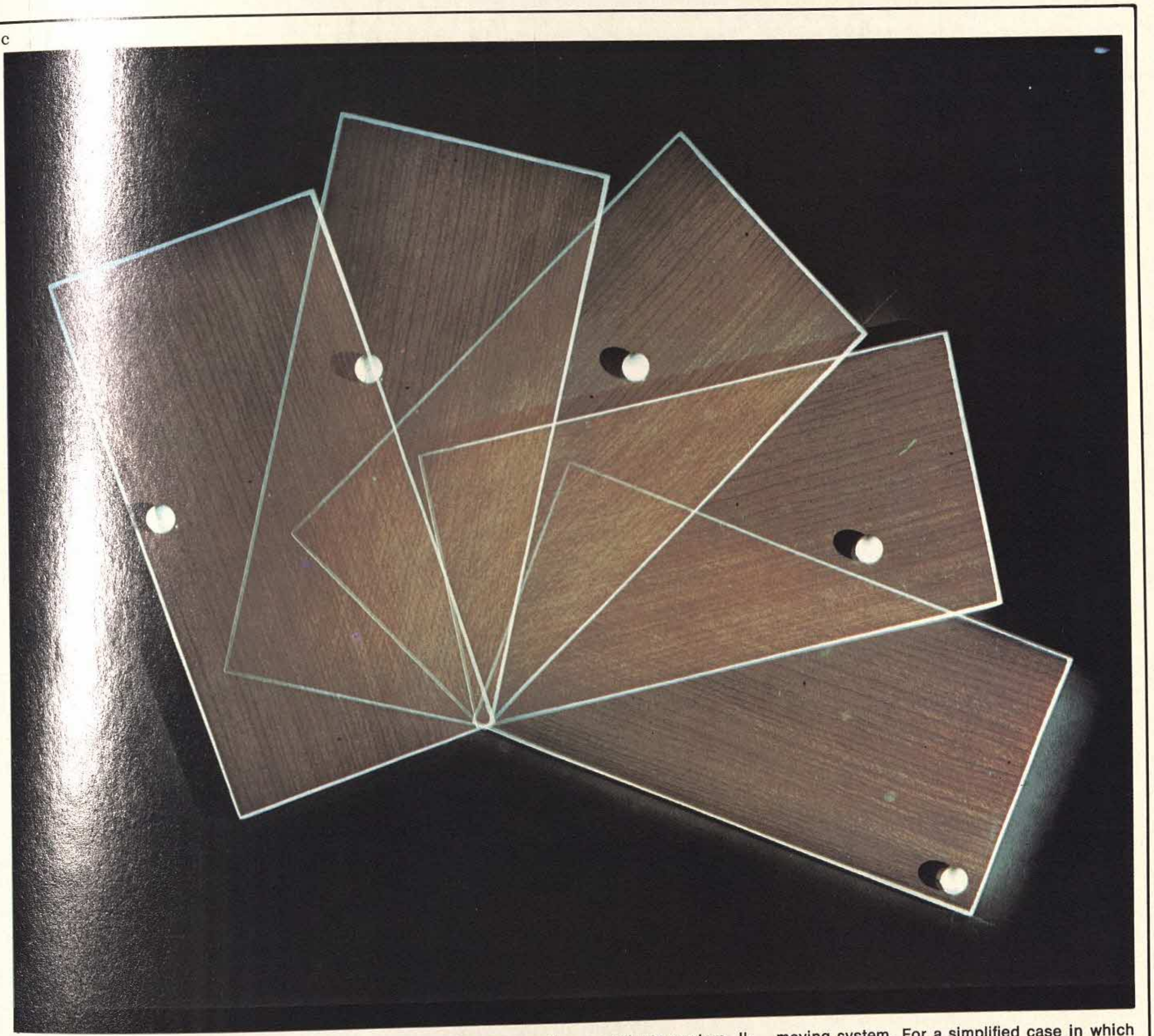


THE CORIOLIS RELATION—A body is again placed on the supporting board. (To keep matters simple, the system used in the previous illustrations representing plane motions rather than spatial ones has been adopted.) This body is accelerated with respect to the board (Illustration 5a). The board is in turn accelerated with respect to the plane of the table on which it is supported. Moreover, the motion with which the board is displaced on the table is both translatory and rotatory. (In this case, as before, $\vec{A}$ is the absolute acceleration; $\vec{a}$, the relative acceleration (the acceleration of the body with respect to the board); and $\vec{A}_t$, the drag acceleration (the acceleration of the board with respect to its supporting plane). In this case, the absolute acceleration is given by $\vec{A} = \vec{a} + \vec{A}_t + \vec{A}_c$.

The term A_c appears only when, as postulated in this case, the board also has a rotatory motion with respect to its support or, in mathematical terms, when the moving reference system has both rotation and translation with respect to the absolute reference system. This is called the complementary acceleration, because it is a term that completes the simple vector sum of the other two.

In a rotating system, the complementary acceleration $\vec{A}_c$ is itself the sum of two accelerations: the centripetal acceleration and the

c



Coriolis acceleration. The Coriolis acceleration is named after the French mathematician Gaspard Gustave de Coriolis, who first described it.

The centripetal acceleration has the mathematical form $\vec{\omega} \wedge \vec{\omega} \wedge \vec{r}$, where ω is the vector of the rotation velocity of the moving reference system, $\wedge$ is the symbol for vector product, and $\vec{r}$ is the radius vector between the origin of the fixed coordinate system and the body whose motion is under study.

The Coriolis acceleration has the form $2\vec{\omega} \wedge \vec{v}$ where v is the velocity of the body with respect to the rotating coordinate system. Illustration 5b shows the direction of the Coriolis acceleration relative to the velocity and rotation vectors.

The force exerted on a body is equal to the product of the acceleration of the body and its mass. In the stationary system the force acting on the body under study can be written as: $\vec{F} = m\vec{A}$, where m is the mass and $\vec{A}$ is the absolute acceleration. Similarly the acceleration in the moving system can be expressed in terms of an effective force: $\vec{F}_{eff} = \overrightarrow{ma}$, where a is the acceleration relative to the moving system. For a simplified case in which there is no translational acceleration ($\vec{A}_t = 0$), $\vec{F}_{eff} = \vec{F} - m\omega \wedge \omega r - 2m\omega \wedge v$. The differences between the real force $\vec{F}$ and the effective force $\vec{F}_{eff}$ are often described as pseudoforces. These pseudoforces arise through the choice of a particular coordinate system to describe the motion and do not follow Newton's principle of action and reaction.

Illustration 5c shows an example of a moving body in a coordinate system that is rotating without any translational motion.

DISPLACEMENT AND MOTION

the shortest way is not always the easiest

An important and fundamental part of physics is mechanics—the branch of science concerned with the study of the motion of bodies and with prediction of this motion from the causes that produce it. For example, mechanics seeks to establish how much thrust must be applied to a space capsule to put it into orbit; how the blades of a gas turbine will vibrate when they are subjected to the violent blast of the jet; how the pendulum of a high-precision chronometer will oscillate when it is delicately pushed by the mechanism of its movement; and what degree of banking must be built into the curves of an automobile race track.

Studies in mechanics presuppose (1) a knowledge of how a body with a certain mass changes its state of motion when a force is applied to it and (2) an understanding of the ways of determining the shape of the trajectory that a moving body will describe. The science of mechanics has an infinity of scientific and practical applications, but these applications all depend on an understanding of the nature of motion and its simpler geometric laws.

DISPLACEMENT AND MOTION

If, during the course of time, the position of a body remains the same with respect to a particular observer, that body is said to be at rest—or motionless—with respect to the observer. The body does not become displaced; it does not undergo any movement. The position of a body can be described by stating where it is to be found with respect to an observer.

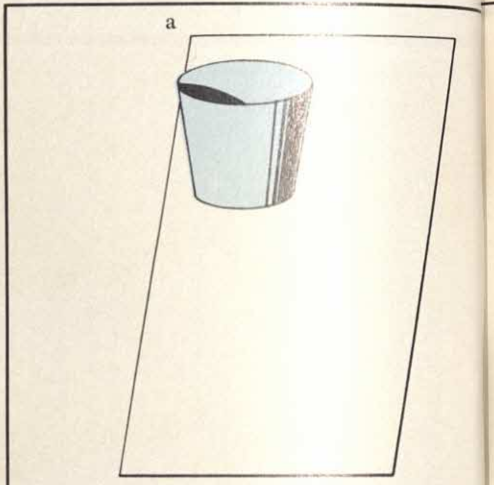
On the other hand, a body may be observed in a certain position at one moment in time and in another position at a later time. For example, an automobile may be seen at a certain location in a square, but when the observation is repeated after the passage of a certain interval of time, the same car may be found at some other location in the square. In this case the car, or the body, has undergone "displacement" (more precisely, the term is used to designate the orientated line segment that joins the first point to the second). Thus, the car has changed position in the course of time, but the path it has followed in becoming displaced is not of interest. The sole interest lies in knowing where the car was at the beginning and where it was at the end. It could have moved along the straight line joining the two end points of its displacement; it could have followed some tortuous path; it could even have left the square and then come back. But all of this is completely irrelevant to the definition of displacement: the displacement is only the change of position that has occurred between the first moment and the last.

Motion, on the other hand, is something quite different, taking into account the trajectory described by the body in arriving at its final position. If the law governing the motion of a body is known, the path it will follow can be deduced. In particular, the straight line joining the initial point to the final point of the displacement is the shortest possible trajectory for any kind of motion capable of producing this displacement.

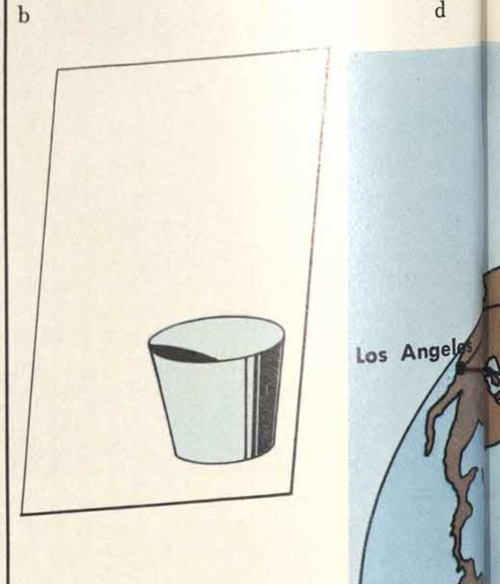
REPRESENTATION OF MOTION

Describing the motion of a body means describing the geometric form of its trajectory, and this depends largely on the factors that produce the motion. Study of the trajectories of moving bodies is, therefore, of great interest in all fields of science and technology. To predict the motion of a body that is subject to known forces, a scientist or engineer applies the laws of mechanics in order to find the trajectory that the body will have to follow. For example, suppose that he wants to know the trajectory, in a vertical plane, of the center of gravity of an automobile in motion. Assume that this vehicle is traveling along a straight road and suddenly hits an object with one of its wheels. If this automobile has the normal configuration of pneumatic tires, spring suspension, and shock absorbers, it will not make a sharp jump but will tend to rise gradually. Before car manufacturers put a ve-

1

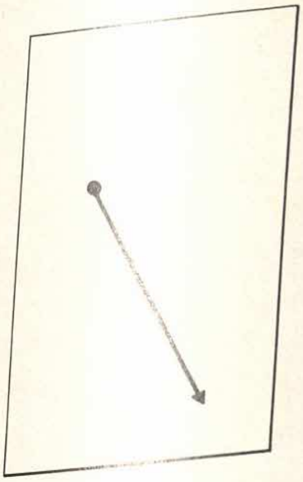


DISPLACEMENT—An object placed on a table occupies a certain position at one moment in time (as in Illustration 1a). Assume that at some later time it occupies another position, as in Illustration 1b. The trajectory it may have followed in moving from the first position to the second is not known. What is known is that the object underwent displacement; its position changed. Since the trajectory the

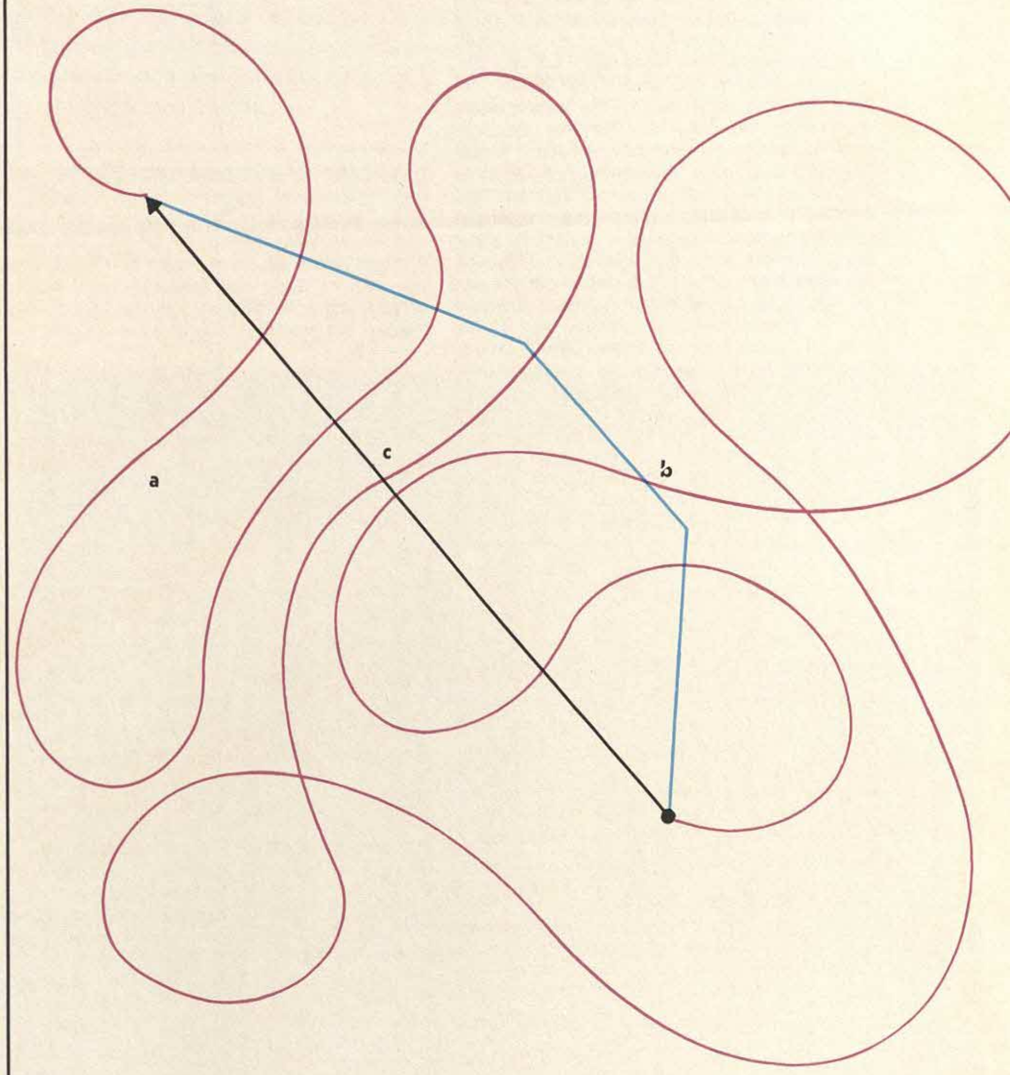


body followed is unknown, it has been deliberately ignored. The displacement can be represented by an arrow that starts at the original position of the body and terminates at its final position (Illustration 1c). The same is true of an airplane traveling from Los Angeles to Rome (Illustration 1d). The course of the airplane, or its various altitudes, are factors of its motion, not its displacement.

c



2



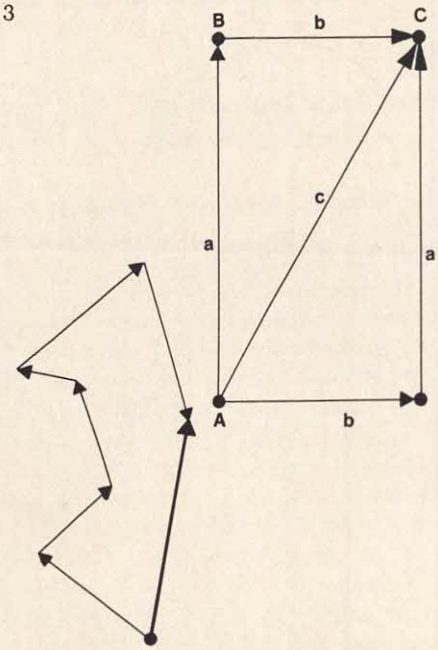
MOTION—The trajectory of a body is represented by all of the points it comes to occupy while it is being displaced from its initial position to its final position. In this illustration, the arrow indicates the displacement of a point of the body, which, in this particular case, is rather large. The same displacement can be achieved by means of an infinite number of motions, ranging from complicated trajectories (red line or **a**), through simpler ones (blue line or **b**), to the simplest of all—the trajectory represented by the arrow that indicates the displacement **c.** The motion of a body from one point to another is determined by the forces that act on it. A displacement can be described by specifying no more than the initial and the final points occupied by the body; to describe a motion, on the other hand, it is necessary also to specify the position of all the intermediate points that the body occupies at successive moments of time.

hicle into production, they study the suspension by carrying out calculations (that sometimes become very complicated) designed to determine the trajectory that the car will follow when it hits certain kinds of obstacles. The form of the trajectory must be known in order to ascertain whether it will cause discomfort or danger to the driver. The designer of the vehicle tries to design the various elements of the suspension system so as to obtain a trajectory that will allow comfort and safety to be maintained. Trajectories that cause discomfort are modified.

In other situations, the opposite procedure may be followed—that is, the geometric form of the trajectory may be studied in order to determine the forces that are acting on the body. For example, an artificial satellite revolving around the Earth at a comparatively low altitude will not follow a perfectly elliptic orbit; rather,

THE COMPOSITION OF DISPLACEMENTS— A body moving from point **A** to point **B** carries out displacement **a.** Subsequently, moving from point **B** to point **C** it carries out a second displacement **b.** This succession of two displacements takes the body from point **A** to point **C** and corresponds to the displacement marked by the arrow **c.** This last displacement is said to be the "sum" of the two displacements **a** and **b.** Interestingly, if the body first carries out a displacement equal to and parallel to **b** (that is, **b′**) and then a displacement equal to and parallel to **a** (that is, **a′**), it will still arrive at the same point, **C.** This simple observation affirms that displacements can be added in any order whatsoever: the sum of the displacements is independent of the order in which they are added. When a body is subject to a succession of displacements, the overall displacement can thus be determined quite easily by drawing all the arrows representing the individual displacements in such a way that the tail of one arrow will coincide with the point of the preceding one. The sum of the displacements will then be represented by an arrow whose tail is at the starting point and whose point coincides with the point of the last arrow in our series. This arrow is also called the resultant.

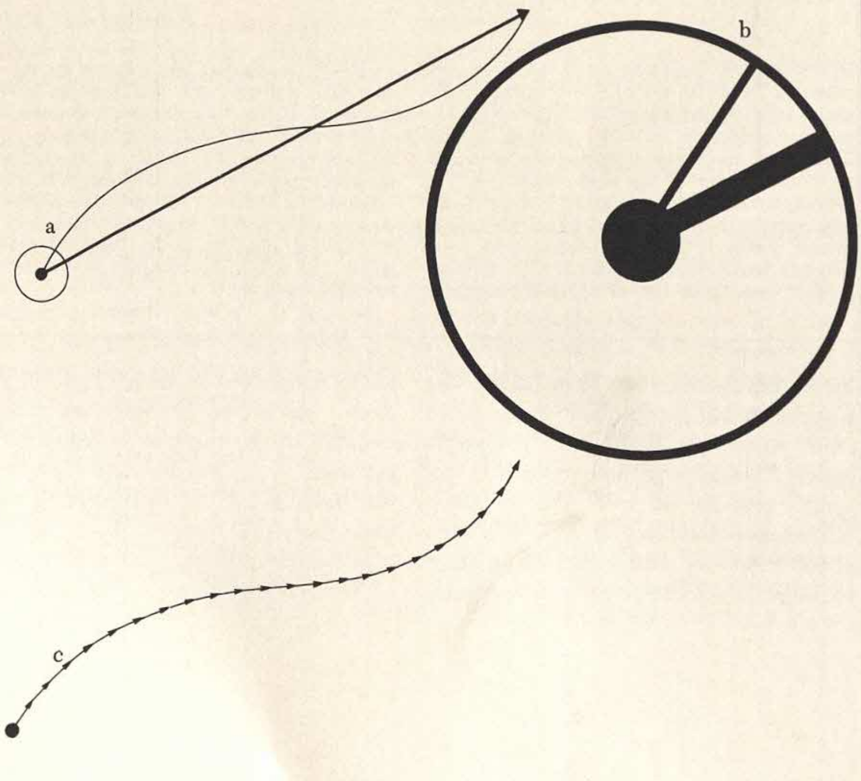
3



A DISPLACEMENT THE SUM OF MANY SMALL DISPLACEMENTS—Diagram 4a represents a complex trajectory of a body that has moved from one position to another. The displacement corresponding to this motion can be represented by means of the arrow that joins the initial point of the trajectory to the final one. A rather precise idea is thus obtained of the difference between the displacement between two points and the trajectory followed in moving from one point to the other.

Diagram 4b shows that, if a small segment of the curve of the trajectory is greatly enlarged, the curvature will be virtually undiscernible and this piece of the trajectory will appear to be a segment of a straight line. Consequently, the whole of the trajectory of the motion can be replaced by a succession of small arrows, each of which represents a very short segment of the trajectory (Illustration 4c). Because segments as short as these are practically straight lines they can readily be equated with the corresponding displacements. By definition, if the trajectory of the motion is considered to be the sum of these displacements, then the overall displacement is also the sum of these small displacements. A single curved trajectory has thus been replaced by a sequence of infinitesimal straight-line displacements.

4



MOTION WITH ZERO DISPLACEMENT—A pendulum oscillating to and fro around a pole of a gyroscope carrying out a precession motion (illustrated in the photograph) are just two examples of a type of motion in which a body periodically returns to a position that it occupied at an earlier moment in time. The time needed for the pendulum or the gyroscopic

5

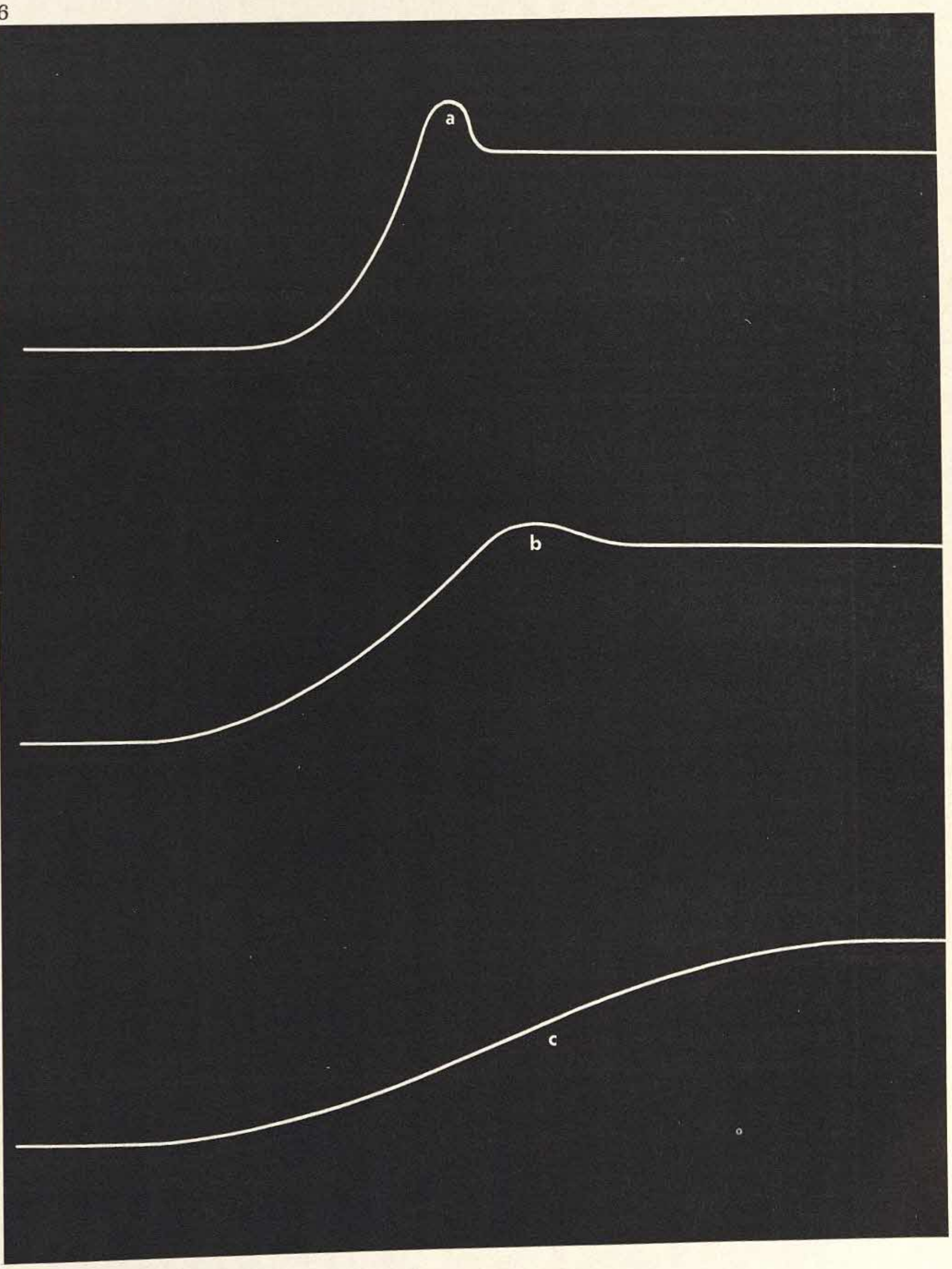
it will deviate slightly because it is subject to the gravitational pull of the continental landmasses. Knowledge of the deviations from the elliptical orbit—that is, of the exact geometric form of the satellite's trajectory—makes it possible to deduce important information about these continental landmasses. For example, observed data may reveal significant facts concerning the distribution of mass within the planet. Such new information would be useful in many fields.

These two examples illustrate the importance of the study of the trajectories of moving bodies; nevertheless, they provide only a slight idea of the very numerous applications of this branch of mechanics.

pole to return to a previously occupied position (arriving at this position with the same speed and in the same direction and sense) is called the period of the motion. In a periodic motion, therefore, the displacement becomes systematically annulled once in every period; this can also be a definition of periodic motion.

THE TRAJECTORY DETERMINED FROM THE FORCE—A vehicle traveling in a straight line encounters an obstacle in the form of a low step and has to jolt over it. The automobile designer must select the suspension and the shock absorbers that will tend to diminish both the occurrence of a jolt and the rapidity with which it takes place. A vehicle without springs jolts sharply, as represented by the curve in Illustration 6a. However, even moderate springing can considerably reduce this jolt, as shown in Illustration 6b. If the car possessed an ideal suspension system—from the point of view of passenger comfort—the jolt would assume the form of the curve in Illustration 6c.

6



INERTIA | a body's property of opposing motion

Inertia is the way in which one of the most important physical magnitudes—mass—manifests itself. The concept of mass escapes all attempts at detailed analysis because it comes close to the concept of the essence of matter. In fact, mass is connected with the quantity of matter that makes up an object.

According to the theory of relativity, the mass of a body at rest differs from its mass when it is moving at a speed close to that of light. The greater the speed, the greater the mass. At the limit of speed, when the body is moving at the speed of light, the mass becomes infinite. For this reason it is better to speak of mass as the inertial characteristic of a body. This concept of mass was introduced by Galileo when he formulated the equation that relates the force to the mass and the acceleration of a body.

In order to arrive at the construction of his equation $F = ma$ (where force F is equal to the product of the mass m multiplied by the acceleration a), Galileo carried out some experiments that became famous not only for their simplicity, but also because they were the first to demonstrate the essence of the phenomena that had been explained in a rather confused manner by the Aristotelians. Moreover, these experiments were taken as models for demonstrating the signifi-

1

THE INERTIA OF A HEAVY BODY—It would be interesting to perform the experiment shown in Illustration 5 with an extremely heavy body, such as a 30,000-ton ship, but suspension of such a huge object from a cable is a practical impossibility. However, the upward thrust exerted by water permits such an experiment. The illustration shows a boat—no different in principle from an ocean liner—resting on the water. Unmoored and not subject to currents, the boat would move in response to a small force applied to its side; the boat would behave as an isolated body in space. In the same manner, a large ship can be moved by a small tugboat pushing against its side. The velocities involved are small; the resistance of the boat or ship to the applied force is also small.

GALILEO'S EXPERIMENTS—The experiments that convinced Galileo that the acceleration obtained in a body is proportional to the force applied are easy to repeat. A small sphere is rolled down an inclined plane. The length of time it takes makes possible the calculation of the average speed of the sphere, or the acceleration to which it was subject during its descent. The inclined plane should be very smooth, hard, and long—at least 2 m (about 6.5 ft). When the plane is constructed in this way, the time of the descent can be measured easily and accurately. The plane must be capable of being inclined at different

2

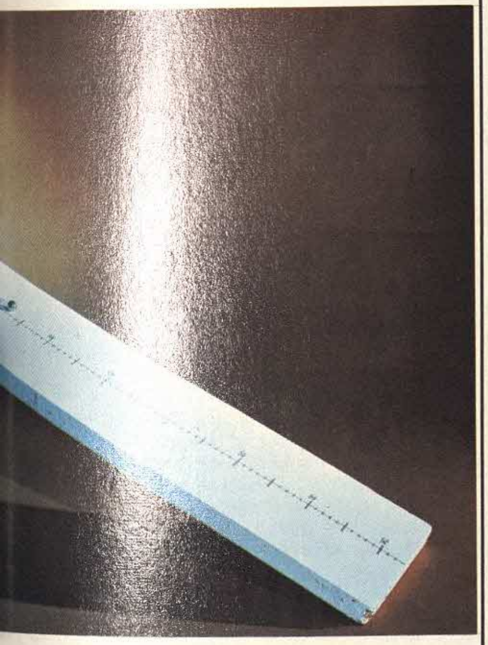
4

a

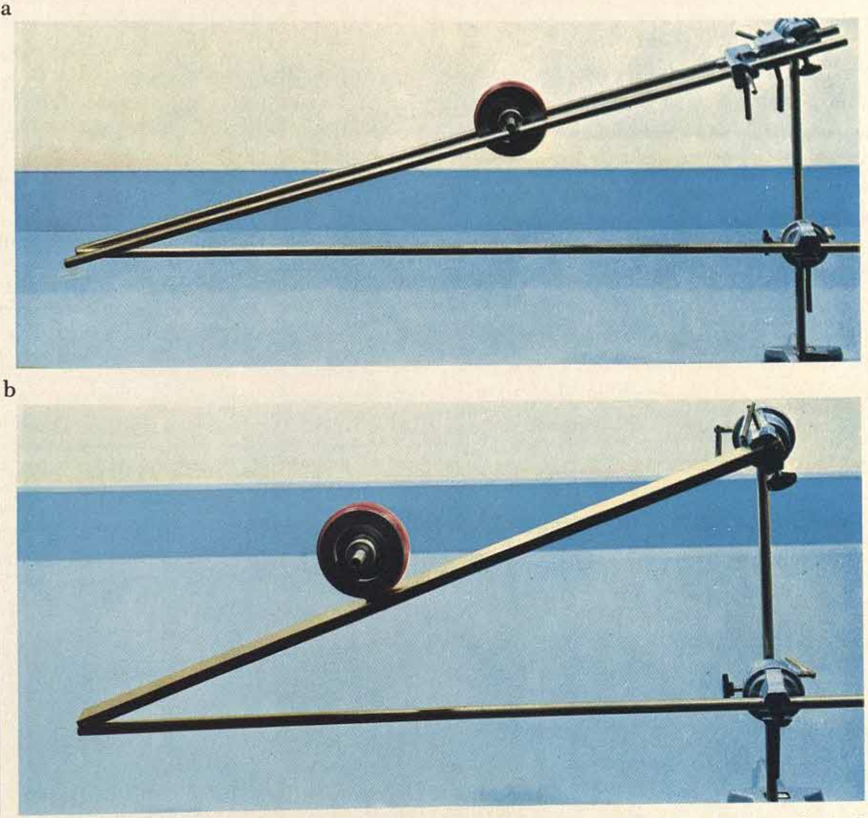
cance of mass and inertia. Even today they are still useful from a didactic point of view.

Galileo decided to examine carefully the special case of freely falling bodies

angles so that different accelerations may be impressed on the sphere. It is not necessary to construct a device that will make it possible to set the plane at any inclination whatsoever. It will be sufficient to have a support whose inclination can be set to certain preselected values that correspond to accelerations (in the sense of its length) that are known fractions of the acceleration due to gravity. The most accurate measurements will be obtained when the plane is long and only slightly inclined; the time of the descent will then be long and the effect of any errors of measurement will be comparatively small.



3



THE INERTIA OF ROTATIONAL MOTION—The fact that a body rolling down an inclined plane is subject to an acceleration that depends only on the inclination of the inclined plane indicates that the body displays a strong inertial characteristic because of its mass, which does not vary from one experiment to another. On the other hand, there are situations in which the body seems to have inertial characteristics different from the ones detected by the indirect method of weighing.

The experiment shown in the illustration involves a disk attached to a concentric axle. Illustration 4a shows the disk rolling down an inclined plane constructed of two rails placed side by side, on which the axle rests. In Illustration 4b, the same disk rolls down an inclined plane on the circumference of its rim. The angle of descent is the same in both experiments; so is the mass of the disk. Nevertheless, the disk rolls down the inclined plane much more rapidly than it does down the rails.

This difference can be explained. In descending the rails, the disk arrives at the end of its run with a small speed of translation of its center and a high angular velocity. In descending the inclined plane, the disk ends its run with a greater speed of translation and a lesser angular velocity. Although the slower descent of the disk down the rails would seem to indicate a greater inertia, it is actually caused by its considerably greater rotational motion and small translation—at times almost zero.

THE STRONG ACCELERATIONS OF A BODY WITH SMALL INERTIA—Strong accelerations are obtained when intense forces are applied to a small mass. An example of this is shown in the illustrations. Illustration 5a shows a brass tube sealed at one end with wax. A plunger has been made from a small rod and a wad of wet cotton attached to a disk, designed to fit closely the interior of the tube. When the piston is slid slowly into the tube—as in Illustration 5b—and then pulled out quickly, the noise of air rushing back into the tube is clearly audible. The mass of air put into motion is equal to the volume of the tube. If this volume is 3 cm^3 (about 0.2 in.^{3}) and the tube has a cross-sectional area of 1 cm^2 (about 0.16 in.^{2}), the experiment will result in 10 cm^3 (about 0.6 in.^{3}) of air—or 0.01 g (about 0.00035 oz)—accelerated by a force of 1 kg (about 2.2 lbs). This acceleration is 100,000 times greater than the acceleration due to gravity. The noise produced is indicative.

to see whether he could make an assumption about their fall that would lead by direct mathematical reasoning to a result that he could verify by experimentation. In finally doing so, he not only provided the foundation for the science of mechanics, which has endured since his time, but also established the fundamental method for the description of experience, which characterizes the whole of modern physics. Galileo's mathematical deduction was a genuine law of motion in the sense that it provided a precise and quantitative description of how freely falling bodies on the surface of the Earth really move.

Galileo introduced also a radical change in the Aristotelian views on violent motion. He developed a strong feeling that horizontal motion on the Earth's surface, unlike vertical motion, is not affected by the Earth and that motion in this direction with any particular velocity should persist unchanged. He verified this by experiments in which a ball was allowed, starting from rest, to roll down one inclined plane and on reaching the bottom was permitted to roll up another plane immediately. The horizontal velocity was retained unchanged no matter how great the horizontal distance. Galileo concluded that the velocity of a moving body that is not under the influence of gravity will not change. This was the important discovery of the law of inertia.

5

MASS AND INERTIA—In everyday life the concept of mass, understood as the quantity of matter, is replaced by the more convenient concept of weight. These are two very different things; mass is a fundamental magnitude, while weight is a force—a mass multiplied by the acceleration due to gravity. The mass represents a characteristic actually possessed by a body. Mass does not vary, regardless of where the body is taken. Weight, on the other hand, varies with the latitude and the height above sea level because the acceleration due to gravity varies in both cases. The weight of a body is the product of its mass multiplied by the acceleration due to gravity. In practice, however, the acceleration due to gravity may be considered more or less equal at all points on the Earth; moreover, the experiments of mechanics are not usually carried out at high altitudes. Therefore, as a first approximation, it may be said that the weight of a body is proportional to its mass

A simple experiment helps put the concept of mass into relief. The experiment is that of attaching a heavy body (and, therefore, a body that has a large mass) to a flexible cable whose upper end is fixed to a solid support such as a beam. The illustration shows an immense log so suspended. If the log were resting on the ground and a man tried to lift it, he would not be able to do so. In physical terms, the attempt to exert a force equal to the weight would have failed. If the same man attempted to lift the log higher than it has been lifted by the cables, he would be no more successful—for the same physical reason.

If, however, the man pushes the log laterally, he can create an appreciable horizontal movement, even if the force he exerts is applied for a short length of time. Lifting the log was impossible because he was unable to exert a force equal to its weight. The weight is the force that the gravitational attraction of the Earth exerts on the log, which puts it into motion with an acceleration of approximately 10 m (about 32 ft) per second per second. In order to lift the log, a man would have to impress the same acceleration in the opposite direction—obviously impossible. Much less resistance was encountered when the man tried to move the log sideways because a much lesser acceleration was required. From this, it can be learned that when a body is freed of its weight by being suspended from a cable, its mass is apparent through its resistance to being pushed—the resistance being proportional to the intensity of the thrust. This proportionality is its inertia—its tendency to resist being put into motion by the action of a force applied from without.

THE MOMENT OF INERTIA

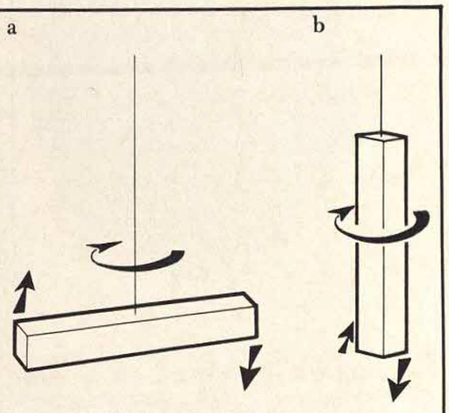
a fundamental concept in the dynamics of rotation

The fundamental laws of dynamics assist in generally explaining the characteristics of the motions of bodies whose masses are known, and the form and forces to which they are subject. However, understanding the motion of bodies in the more complicated cases requires not only a review of these general considerations but also some formal simplifications pertinent to the problems presented. As an example, the calculation of the motion of the celestial bodies would be extremely complicated if Newton had not demonstrated that gravitational forces manifest themselves as if the mass of spherical bodies were concentrated at the center. The study of an entire class of motions, the circular ones, is greatly simplified by the concept of the moment of inertia.

1

RESISTANCE TO ANGULAR ACCELERATION—Illustration 1a shows a long, thick beam suspended by a thin cable that is attached to its center of gravity. The thin cable twists very easily. If the beam has a considerable mass, it will rotate rather slowly, even when maximum force is applied to the ends.

In Illustration 1b the beam is suspended by one of its ends so that it hangs vertically. When a couple is applied to it, a much greater effect is obtained than in the preceding example. Now, little effort will make the beam rotate quite rapidly. It is as if the beam under these conditions exerted less resistance to being put into motion. This phenomenon is explained by the moment of inertia.

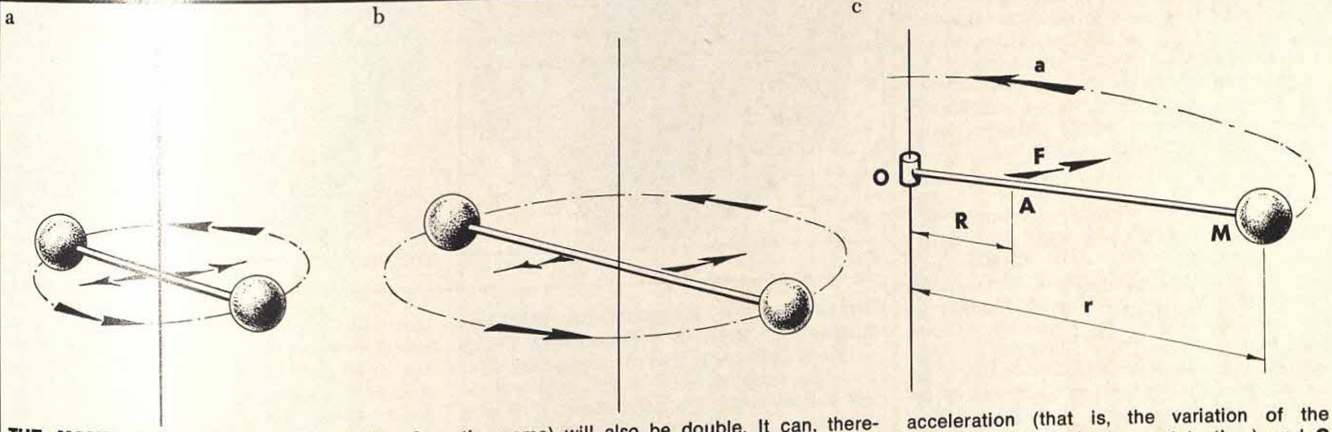


When a body moves in a straight line under the action of a force, the second law of dynamics requires that its acceleration (the variation of the velocity caused by the fact that a force is acting on the body) be proportional to the intensity of the force that is acting and inversely proportional to the mass of the body.

By using a certain force to accelerate different bodies, accelerations are obtained that are inversely proportional to the mass of the body in each particular case. The mass, therefore, is the factor opposing a strong acceleration in rectilinear motion.

If, on the other hand, a circular motion is considered, it is convenient to replace the concept of a single force with that of a couple of forces, because a couple always acts to initiate the rotation of a

2



THE MOMENT OF INERTIA—Illustration 2a shows two heavy spheres connected by a rigid rod that is practically weightless. The spheres are made to rotate about the center point of the rod. For a rotation velocity of 1 rps (revolution per second) a certain effort must be exerted.

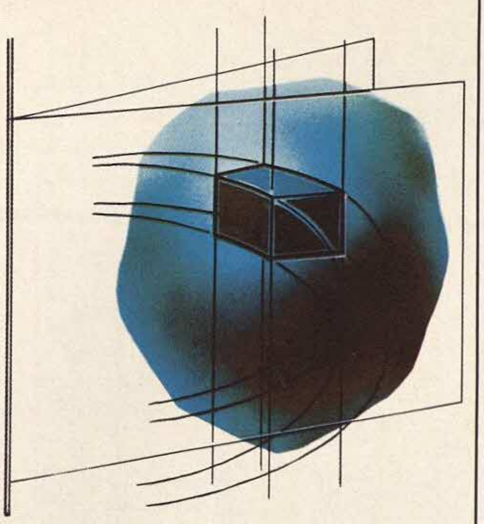
Illustration 2b shows two spheres identical with those used in Illustration 2a, but placed at twice the distance apart. In this case, more resistance is met in trying to achieve the same rotational velocity. The reason for this difference is that the extended spheres have to cover twice the distance of those in Illustration 2a before they make a complete turn; therefore, their velocity (the linear velocity, of course, because the angular velocity remains the same) will also be double. It can, therefore, be said that the second system has a greater moment of inertia than that of the first.

The moment of inertia of a body can be defined with the aid of Illustration 2c, which shows a light bar, hinged at one of its ends **O,** along which a weight **M** may be moved. (The mass of the weight is much greater than that of the bar, so that the mass of the bar may be considered negligible.) Consider the application of a force **F** at point **A** at a distance **R** from the hinge **O.** Let **r** represent the distance between point **O** and the point at which the weight **M** is fixed. If **R** and **F** remain constant, the angular acceleration impressed on the bar will become greater as the distance **r** is reduced. If α is the angular acceleration (that is, the variation of the angular velocity with respect to time) and **C** is the moment of the force **F** about the center of rotation **O,** it can be demonstrated that these variables are connected by the relationship $\alpha = C/I$. The angular acceleration is directly proportional to the moment the force exerts on the system and inversely proportional to a magnitude **I,** which in this case is given by $I = Mr^2$. This equation becomes identical with the one that expresses the second law of dynamics ($a = F/M$) when the moment of the force is replaced by the force itself, the angular acceleration is replaced by the linear acceleration, and the magnitude **I** (the moment of inertia) is replaced by the mass.

body. Moreover, not only is it the mass that inhibits the body from beginning to rotate, but also the distribution of the mass around its axis of rotation. The moment of inertia is a magnitude that takes into account this distribution of the mass, making it possible to predict the accelerations that will occur in rotational motion.

This concept is extremely important in applied mechanics because it plays a part in all types of rotational motion in which variations of the angular velocity occur. The intensity of the accelerations that can occur in this kind of motion depends on the moment of inertia of the body concerned. For example, the length of time the gas turbine of a jet engine takes to reach its operating speed depends on the moment of inertia of the rotor. If the rotor is to be capable of accelerating rapidly, it is essential that its moment of inertia be low.

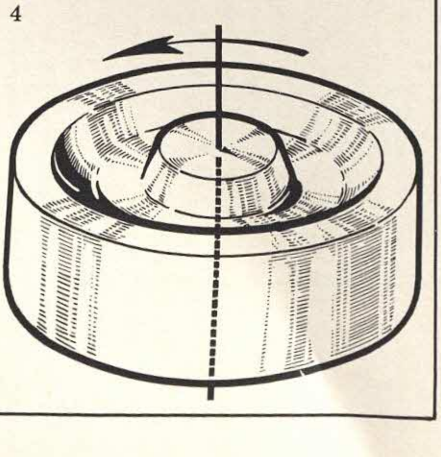
3



THE MOMENT OF INERTIA OF A LARGE BODY—The case of the moment of inertia of a simple body finds rare applications. A frequent case, however, is that in which the moment of inertia of a large body is calculated. It is important to understand that the moment of inertia of a body depends on the position of the axis of rotation. The beam discussed in Illustration 1 had a different moment of inertia, according to the axis around which it was made to rotate. In general, therefore, the principle on which the evaluation of the moment of inertia is based consists of breaking up the body into many elements of small dimensions—so that all the points of each element are at the same distance from the axis of rotation. The moments of inertia of all these elements are then added together. The illustration shows the principle on which this decomposition is effected in the case of a body of irregular shape. If the shape of the body is not completely irregular, the calculation of the moment of inertia can also be carried out by means of integral calculus, avoiding laborious numerical summation.

MOMENT OF INERTIA AND KINETIC ENERGY—A body of cylindrical form rotating about its axis of symmetry is shown. Although the center of gravity of the body is situated on the axis of rotation (and, therefore, remains stationary), the body still possesses kinetic energy by virtue of its rotational motion. The formula that enables the energy content of the system to be evaluated is identical with that for the kinetic energy of purely translatory motion, if the moment of inertia **I** is replaced by the mass **m** of the body, and the angular velocity ω by the linear velocity **v.** The formula is $\mathbf{E} = \mathbf{I}\omega^2/2$.

4



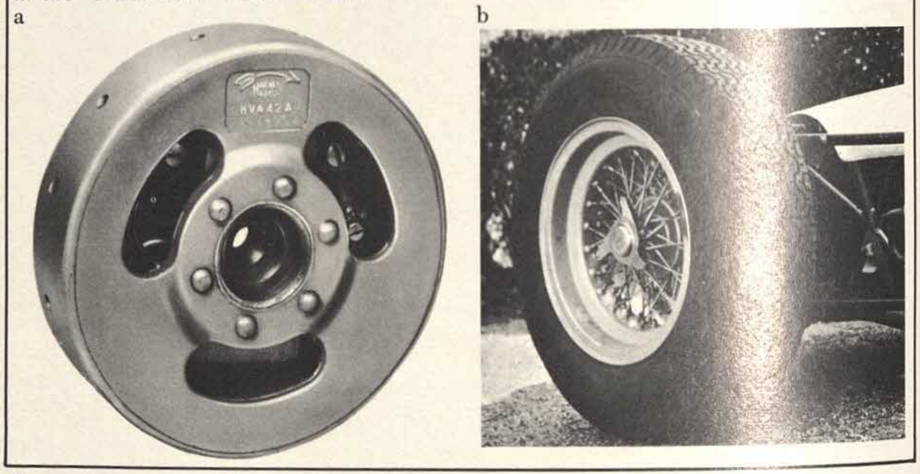
5

THE FLYWHEEL—A rotating body always possesses kinetic energy, even when its center of gravity is fixed. If the body rotates at a high speed and has a large moment of inertia, it will have considerable kinetic energy.

The flywheel is a mechanism that consists of a body with a large moment of inertia. It is used to steady the speed of rotation of a system in which the motion is produced by a couple that is not perfectly constant. Single-cylinder internal-combustion engines need flywheels with a considerable moment of inertia to regularize the rotation; in the absence of such a flywheel, the motion would be highly irregular, especially in the case of four-stroke engines in which energy is supplied only once in two revolutions of the crankshaft.

Illustration 5a shows the flywheel of a motorcycle engine.

Sometimes a part of the system itself will act as a flywheel, with harmful effects on the overall functioning of the system. For example, the wheels of an automobile (Illustration 5b) act as flywheels. Their rotation absorbs a great deal of energy that could be much better employed in accelerating the vehicle and must eventually be reabsorbed by the brakes when the vehicle must be halted. In racing cars, the moment of inertia of the wheels is reduced to a minimum by reducing the mass of the disks. The wheel shown here is used in competition racing; it is lighter than a full disk.

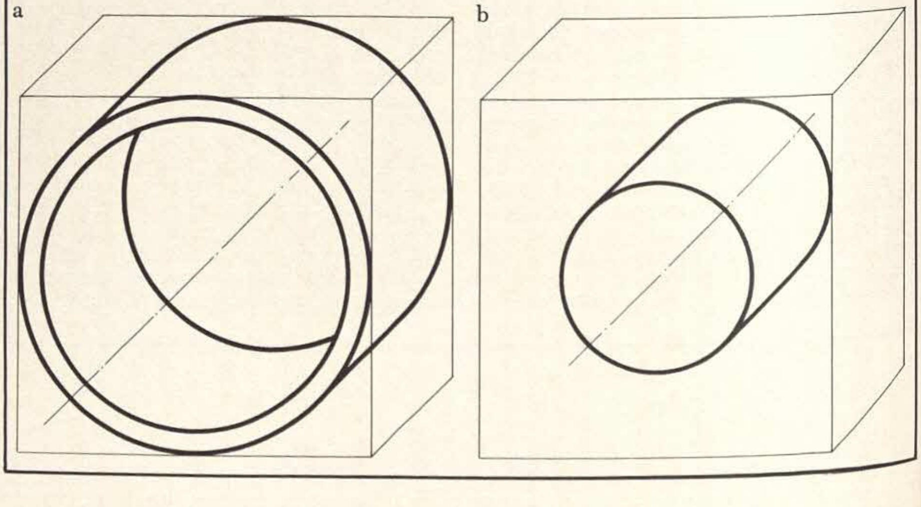


a

b

6

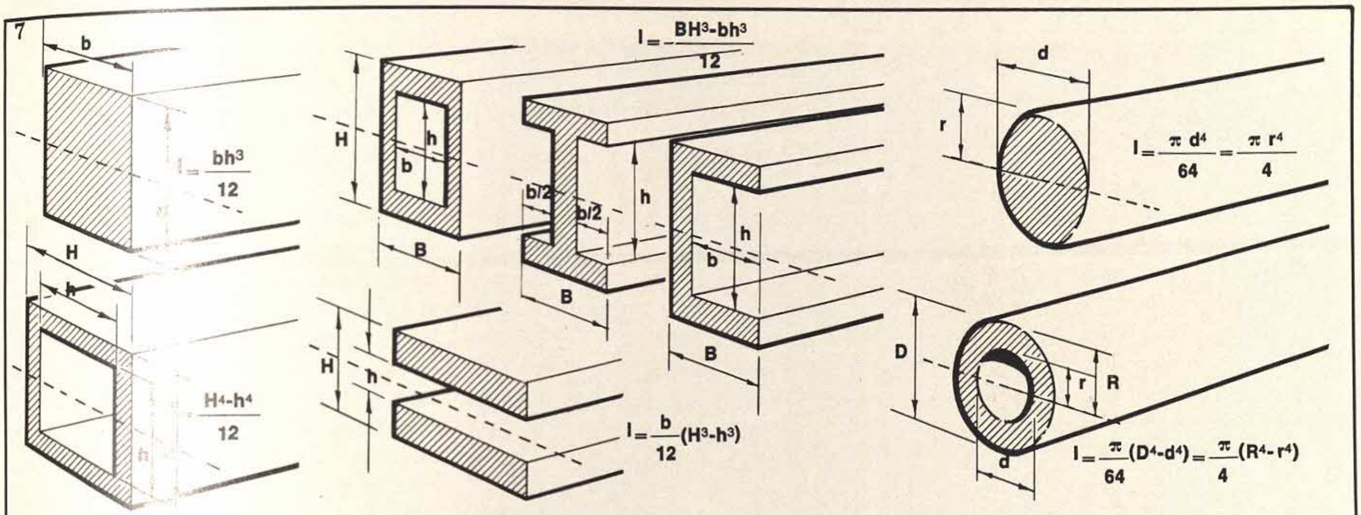
OBTAINING THE MAXIMUM OR MINIMUM MOMENT OF INERTIA WITH A GIVEN MASS—This experiment is that of constructing from a certain mass of metal a body that will rotate entirely within the cube shown. (The axis of rotation is indicated by the dotted line.) In Illustration 6a, the solution is that in which the body has the maximum moment of inertia; Illustration 6b shows the solution with the minimum moment of inertia. In the first case, the mass had to be distributed around the axis of rotation in such a way as to maintain the maximum possible distance. In the second case, the mass has been placed as close as possible to the axis of rotation.



a

b

7

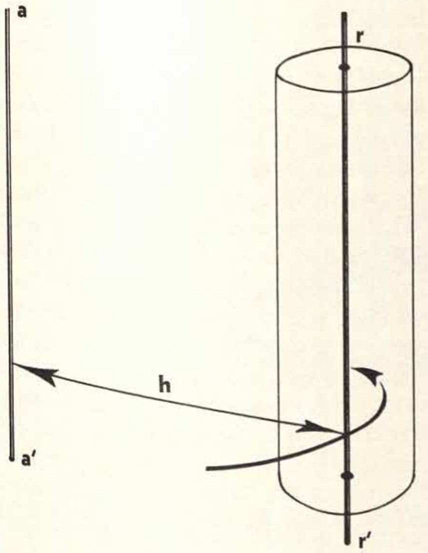

b
$I = \frac{bh^3}{12}$
H
h
$I = \frac{H^4-h^4}{12}$
h
$I = \frac{BH^3-bh^3}{12}$
H
h
b
B
b/2
b/2
h
B
h
b
B
H
h
$I = \frac{b}{12}(H^3-h^3)$
d
r
$I = \frac{\pi d^4}{64} = \frac{\pi r^4}{4}$
D
r
R
d
$I = \frac{\pi}{64}(D^4-d^4) = \frac{\pi}{4}(R^4-r^4)$

SOME MOMENTS OF INERTIA—This illustration shows the value of the moment of inertia about the geometric axis (an axis of symmetry) of a number of plane surfaces commonly used in mechanical constructions. As can be seen from the formulas, the geometric moment of inertia (per unit length of the object) has the dimensions of a length raised to the fourth power. This is because the moment is obtained by summing the products of the various elements of area (which is proportional to the mass per unit length), multiplied by the squares of their respective distances from the axis (indicated by the dotted line) to which the moment of inertia is referred. These formulas are of great importance in the statics of elastic structures (see Illustration 9).

THE THEOREM OF PARALLEL AXES—The calculation of the moment of inertia of the cylindrical body shown in this illustration with respect to the axis **aa′** would be easy if the axis of rotation were also the axis of symmetry. However, this is not the case.

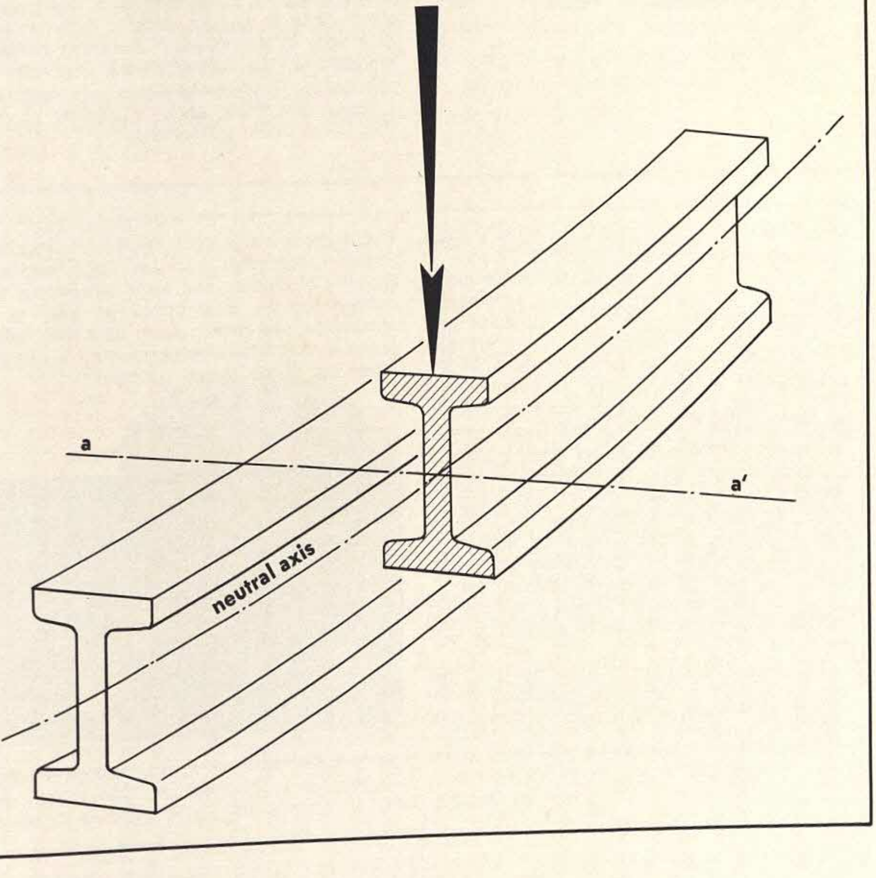
Breaking up the body into small elements (in accordance with the method shown in Illustration 4) would lead to a very complex calculation. A theorem demonstrated by the French mathematician Joseph Louis Lagrange in 1783 is helpful. This theorem states that the moment of inertia **I′** with respect to the axis **aa′** is equal to the moment of inertia **I** that the body has with respect to the axis **rr′** (which passes through its center of gravity and is parallel to aa′), plus the moment of inertia that the mass of the body possesses with respect to the axis **aa′** when the mass is considered as concentrated at the body's center of gravity. Therefore, $I' = I + mh^2$.

8



AN APPLICATION TO STATICS—The bending of a beam supported at both ends occurs in such a way that the longitudinal section—the so-called neutral axis (indicated by the dotted line)—will bend without becoming either lengthened or shortened. Any section lying above the neutral axis will become shortened, whereas the sections below it will lengthen. The wider the section of the beam farthest away from the neutral axis, the better will the beam resist the applied bending moment. The reason for this is that the beam is strongest at the very places where it is subject to the greatest stresses because of the lengthening or shortening. If the bending resistance of a beam is calculated as a function of its form, its resistance will be inversely proportional to the moment of inertia of the beam with respect to an axis **aa′**—that is, perpendicular to both the neutral axis and the direction of the bending.

9



IMPULSE AND MOMENTUM

two basic quantities of mechanics

When a moon rocket thunders away from its launch pad, it is indeed an awesome sight. Vast clouds of smoke and flame billow up as the rocket ascends, creating the dramatic "fireworks" that provide spectacular evidence of the force, or thrust, propelling the rocket.

Rocket thrust is a precisely calculated force, measured in terms of *impulse,* which is one of the basic concepts of mechanics. Impulse is related to the force applied to a body and the time through which the force acts; numerically, it is the product of a force and the time in which the force acts.

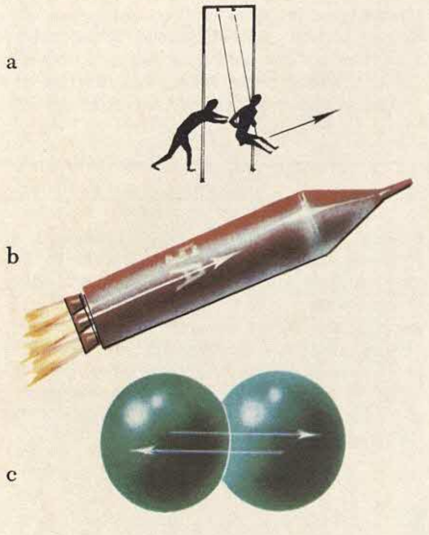
Newton's second law of motion states that applying a force to a body causes its displacement, producing an acceleration that is proportional to the force and inversely proportional to the mass of the body. If the mass of the body, the force, and the amount of time for which it has been applied are known, then the velocity of the body and the distance through which it has traveled can be calculated. This calculation can sometimes be done more simply by the concept of impulse.

Another very useful concept in mechanics—a concept with such a vast range of applications that it is used both by astronomers and nuclear physicists—is the quantity of motion called linear momentum. (The momentum of a body is equal to the mass of the body times its velocity.) Impulse and momentum are related to each other. If an impulse Ft is given to a body of mass m during a time t, the body will undergo an acceleration a. Therefore, $Ft = mat$ because, according to the second law of motion, $F = ma$. However, a body initially at rest, if subjected to a constant acceleration a during a time t, will attain a velocity $v = at$; thus, $Ft = mv$. The impulse is, therefore, equal to the change in the momentum of the body.

CONSERVATION OF LINEAR MOMENTUM AND THE THIRD LAW

Suppose a number of spherical bodies are launched into space. Each will move in a different direction at its own velocity. While these spheres are in motion—motion that has been derived from the initial impulse—a number of mechanical phenomena may occur. For example, the bodies may attract or repel each other (because of electrical forces) and collide. Whatever happens, the total momentum of the system made up by all these bodies will be conserved—that is, the total momentum will not vary in time. Experiments confirm that when two bodies interact—by attraction, repulsion, collision, or any other kind of reciprocal action—the sum of the momenta possessed by each body before interaction is equal to the sum of the momenta each possesses after interaction.

1



THE EFFECT OF IMPULSE—Impulse is the product of a force and the length of time the force is applied to a body. Numerous examples exist where the force on a body remains constant for the whole time of its application. For example, if somebody pushes on the seat of a swing on which another person is sitting, the push produces an impulse (Illustration 1a). A rocket engine driving a space vehicle for a certain time is exerting an impulse (Illustration 1b). A billiard ball hitting another exerts an impulse for the length of time in which the two remain in contact (Illustration 1c). Impulse is expressed in dyne-sec, newton-sec, and so forth.

Impulse can also be defined when a force of variable intensity is applied to a body. Here the impulse imparted by the force is divided into a number of successive short intervals of time, on the approximation that the force remains constant over each short time interval.

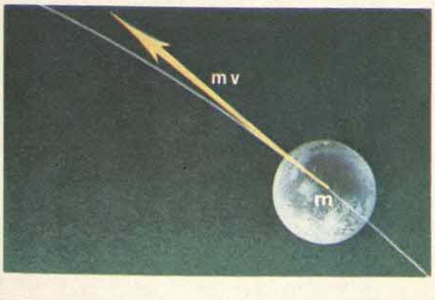
The effect of an impulse on a freely moving body is to vary its velocity. If the motion of a body is restrained, it will either become immobile or the restriction will condition its movement. To understand this, it is important to know something about a quantity of motion called linear momentum.

THE QUANTITY OF MOTION—A body moving uniformly, with a mass *m* and a velocity *v*, possesses a quantity of motion called linear momentum (momentum, for short). Momentum is defined as the product of its mass and velocity: *mv*. The velocity of the body can be represented as an arrow pointing in the direction of movement. The momentum can be represented by the same arrow drawn to a length proportional to the magnitude of the momentum. If the body is not moving along a uniform straight line, its momentum is expressed as the product of its mass and its instantaneous velocity. For example, the momentum of the moon, instant by instant, is equal to the product of its velocity and mass, and is represented, as shown in Illustration 2, by an arrow at a tangent to the lunar orbit at the point which the moon is passing at that moment.

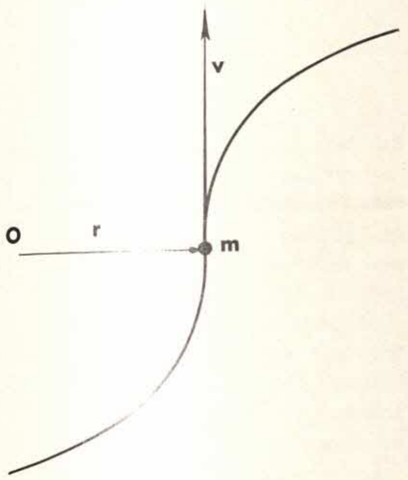
Momentum is a quantity of motion that can be represented by an arrow indicating its intensity and direction and path of movement.

The same consideration applies to impulse. Physical quantities such as force, impulse, quantity of motion, and many others that are represented by a direction as well as a magnitude, are called vector quantities, while physical quantities represented by a number alone are called scalar quantities.

2

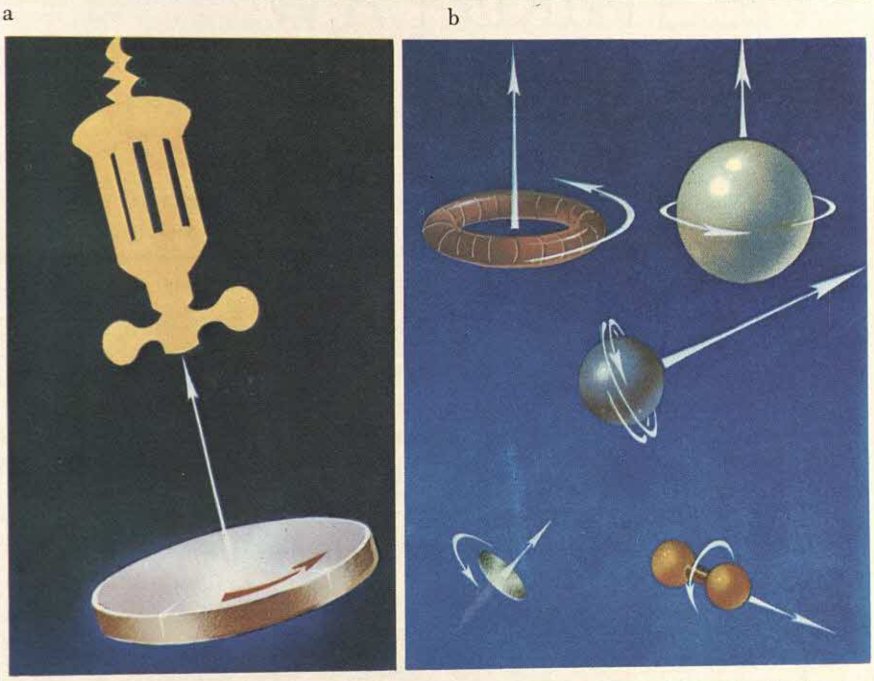


3



ANGULAR MOMENTUM—A body with a mass *m* passing near the point *O* has a velocity *v* and, therefore, a momentum *mv* represented by an arrow applied to the body. The velocity of the body causes it to pass at a distance *r* from the point *O*. The angular momentum of the body about *O* is defined as $r \times mv$, a formula somewhat similar to the definition of the moment of a force. Angular momentum is therefore the moment of the momentum of the body with respect to the point *O*. Just as momentum is an important quantity in the study of a freely moving body, angular momentum is an important quantity in the study of a body in rotary motion around an axis or a point.

4

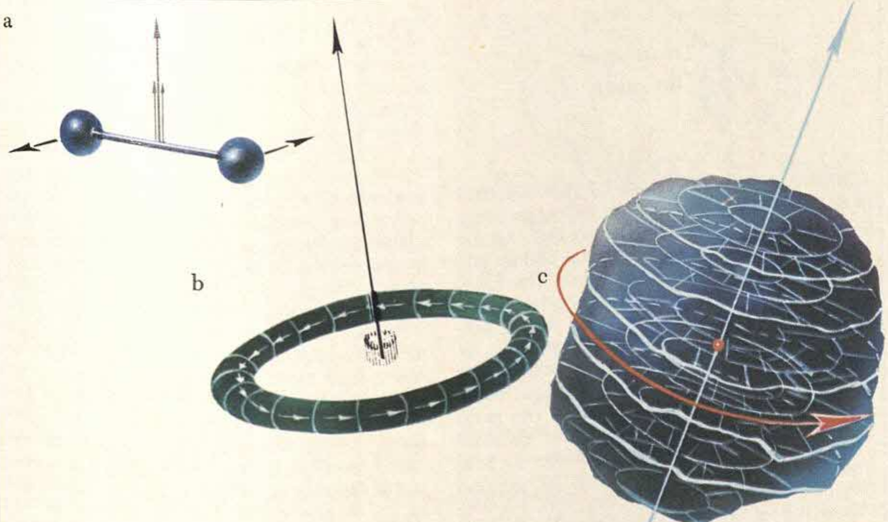


REPRESENTING ANGULAR MOMENTUM—Angular momentum is inherent in the rotation of a body and exists because that body rotates. It is an expression of how fast a body rotates and also the direction of its rotation. Thus, angular momentum is represented by an arrow whose length is proportional to the magnitude of the angular momentum, and whose direction (as shown in Illustration 4a) is that of the progression of the point of a corkscrew.

Illustration 4b shows several objects suspended in space and rotating. Each body has a rotational velocity, mass, dimensions, and a form different from that of the others. An arrow has been drawn around each body; the length of the arrow indicates the amount of its rotational velocity. Along the axis of rotation of each body an arrow has been drawn to represent, by its length, the amount of angular momentum.

From this fact, which can be derived from the second law of motion, it follows that in a collision between two bodies, each will exert on the other an equal and opposite force. This statement corresponds to Newton's third law of motion, which states that all forces arise from the mutual interaction of particles and in every such interaction the force exerted on the one particle by the second is equal and opposite to the force exerted by the second on the first. A law, or principle, is understood as a truth accepted as a valid premise from which to deduce other facts. Because the application of Newton's third law has never led to contradictions, it is accepted as a valid postulate.

5



THE ANGULAR MOMENTUM OF A CONTINUOUS BODY—The angular momentum of more complex bodies than those shown in Illustration 4 can be calculated. Illustration 5a shows a body made up of two masses rotating around the same axis. The total angular momentum is equal to the sum of the angular momenta of each body: the two smaller arrows at the center represent the angular momenta of the two masses, and the longer arrow, twice the length of each small arrow, represents the total angular velocity of the system.

Illustration 5b shows a rotating ring represented as a large number of masses, each possessing an angular momentum. The total angular momentum is the sum of the angular momenta of the small masses and is represented by an arrow equal to the sum of the arrows representing the masses into which the ring is divided.

The angular momentum of an irregularly shaped body (Illustration 5c) is calculated by first dividing it into planes passing through the axis of rotation, then dividing each ring into small masses. The angular momentum is the sum of the angular momenta of all the small masses. While the angular momenta of the masses into which each ring is divided are equal, the angular momenta of the various rings, however, are not equal. The successive rings have different radii and, consequently, different velocities and momenta. The angular momentum of a body may be expressed very conveniently by means of the moment of inertia of the body.

THE THIRD LAW OF MOTION

interacting forces

In a universe so vast that it appears to be limitless, man lives with the assurance that three relatively simple laws govern the motion of matter. These three principles were expressed so well by Isaac Newton late in the seventeenth century that they are still called Newton's laws of motion. The principles enable man to understand and predict the motion of all ordinary bodies.

The first of these laws, known before Newton's time and sometimes called the law of inertia, states that if a body, whether at rest or moving, has no force acting on it, it will remain at rest or will continue to move at a constant velocity in a straight line. The second law states that the product of the mass and the acceleration of a body equals the force that acts on the body. The third law of dynamics, although it has a wide range of practical applications, is more difficult to comprehend. Simply stated, it says that for every action there is an equal and opposite reaction.

THE SIGNIFICANCE OF PUSHING

In discussing certain simple phenomena of mechanics, words often fall short of giving an exact description of reality. However, the third law of motion is clarified by a description of a man who wants to jump into the middle of a road. He plants his feet firmly against the curb and then quickly pushes against it. In this way, he launches himself forward and reaches the middle of the road. The man, who feels the effect of the effort he has made in the muscles of his legs, may say that he jumped because of the pressure of his feet against the curb. This is true, but it is also true that the curb exerted pressure against the jumper's feet. He would not have been able to make his jump had there been no resisting action on the part of the curb.

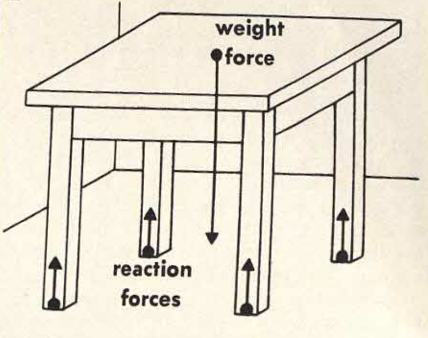
In an imaginary situation, the jumper, immediately after leaving the curb, can remain suspended in the air momentarily. Ignoring the resistance of the air, if he again tries to move his legs in the same way as he did on the sidewalk, he will not increase the length of his jump. Yet moving his legs in midair requires the same expenditure of energy and is just as tiring as before.

It is thus useless to push unless some-

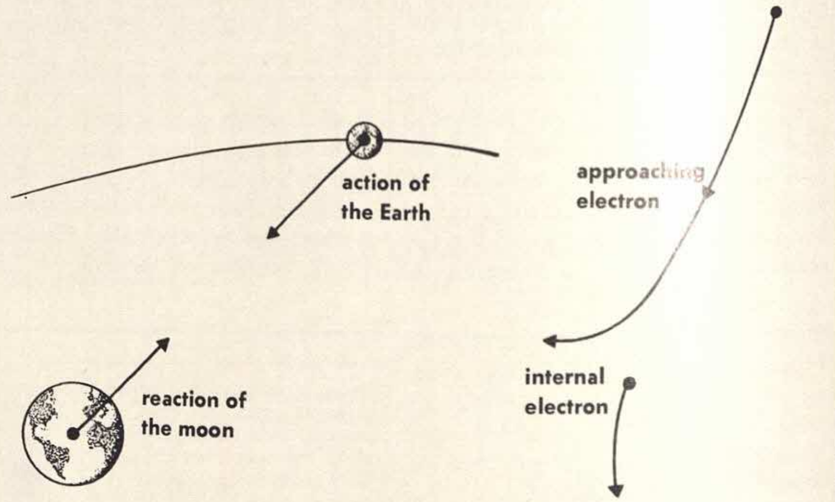
STATIC REACTIONS—The man jumping into the street represents a dynamic illustration of the principle of reaction. The presence of reaction can be observed equally well in a static phenomenon. Each time a force is applied to a body, its velocity changes. If it is at rest, for example, it will begin to move. Within the Earth's gravitational field, all bodies are subject to the effect of their own weight and will fall if free to do so. If they are resting on something, however, they do not move. This is readily apparent by observing all the objects in a room. The fact that they are resting on something and remain still does not mean that they are not subject to weight. A body suspended by a cord will remain still, but the cord will break if it is not strong enough, thus showing that the weight force was active when the body was at rest. It is the weight force that caused the cord to break.

For a body to remain at rest, both the action of the weight and an equal and opposite reaction on the part of the support are necessary. The two forces, acting together, cancel each other. What happens when a book is placed on a table? A weight is added, and the force acting on the table has been increased. The table does not move because the floor on which it is resting reacts with increased force. The heavier the book, the greater the reaction force. One characteristic of all reactions is that they adapt to the actions that produce them.

1



2



ACTIONS AND REACTIONS OF BODIES NOT IN CONTACT—A reaction to a force can occur even in the absence of actual contact. An example is furnished by the celestial bodies and their mutual gravitational attraction. The Earth exerts a certain force of gravitational attraction on the moon, and the moon reacts with an equal force of gravitational attraction on the Earth. If this force were not balanced by the centrifugal force of the moon in its orbit, the moon would fall onto the Earth. The force of gravity exerted between the two celestial bodies is identical. It is not possible to separate the action of the first on the second from that of the second on the first; the two forces are of equal magnitude, but acting in opposite directions.

Another example is that of two charged atomic particles. An electron launched into matter will be forced to traverse the space between the atoms or even inside the atoms themselves and will approach another electron. Because each particle is charged with the same negative force, the two electrons will repel each other. The first electron will repel the internal electron, which will retreat a certain distance from its position. At the same time, the trajectory of the first will deviate. The force that caused these two movements is the same. A reaction has taken place, equal to the original action and inseparable from it.

3

AN IMPORTANT APPLICATION—A man standing in a small boat steps into the stern to jump ashore. He uses the edge of the stern in the same way that the curb was used in the first example. As he jumps ashore, the boat moves away; the force he has exerted on the boat is equal to that exerted on him by the boat. Since the boat moves, he feels less effort in the muscles of his legs and moves over a shorter distance than if the boat had not moved. Even though the boat was free to move, the man was able to make his jump by pushing against it. The same reasoning process applies to the launching of a rocket, as shown in Illustration 4.

On the basis of what has been demonstrated, the following conclusion can be drawn: when a body exerts a force (action) on another body, the second acts on the first with an equal and opposite force (reaction). Regardless of the way in which the two bodies act on each other (contact, electrical, or gravitational force), the principle of interacting forces is always valid.

4

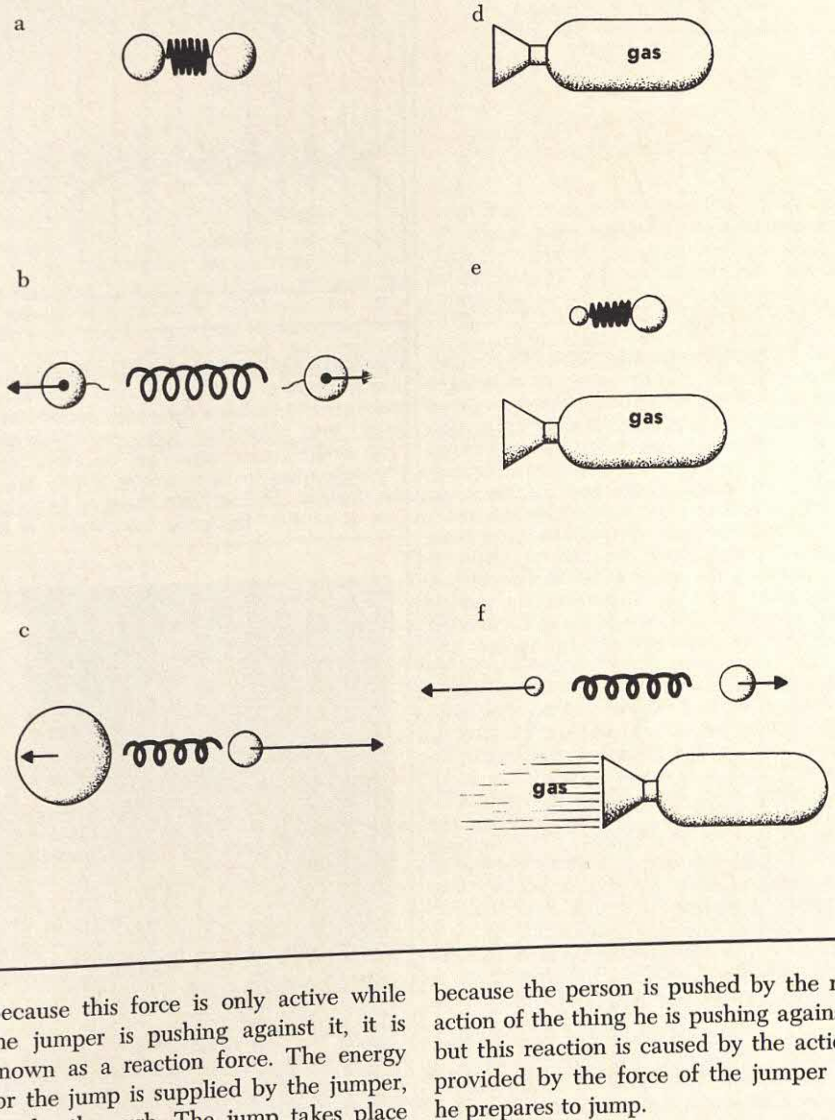
HOW A ROCKET ENGINE WORKS—The functioning of a rocket engine is an important application of the principle of action and reaction (Newton's third law of motion). It can be compared to the action of two identical heavy spheres separated by a spring that is kept compressed by a wire between the spheres (Illustration 4a). If the wire that keeps the spring compressed is cut, the spring will expand (Illustration 4b). The two spheres will be projected in opposite directions at the same velocity, each acting on the other (through the spring) in an identical manner. Because they have the same mass, they will have the same acceleration.

If the experiment is repeated with two spheres of different mass, the reciprocal action will be identical and the two spheres will act on each other with the same force (Illustration 4c). However, because the masses of the two spheres are different, the lighter sphere will be projected at a greater velocity than the heavier.

Illustration 4d represents a rocket engine and shows a cylindrical chamber containing a mass of compressed gas. Initially, the rocket containing the gas is at rest. Upon firing, the gas will expand and rush out the nozzle on the left.

Illustration 4e shows both the rocket engine full of compressed gas and the two spheres separated by a compressed spring. The mass of the two spheres represents the combined masses of the engine and the gas; the spring represents the energy of the compressed gas.

In Illustration 4f, the compressed energy is released by allowing the gas to expand outside the rocket. The action of the escaping gas provokes an opposite reaction in the rocket itself, moving it toward the right. The presence of air impedes a rapid escape of gas and thus diminishes the efficiency of the rocket. Rocket engines function better in a vacuum than in the atmosphere.



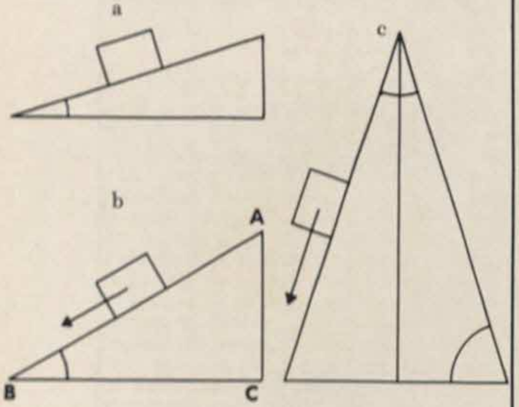
thing is pushed against. The resistance of the thing that is pushed against produces the forward motion. The jumper moves because the surface he was standing on acted on him with a special force. Because this force is only active while the jumper is pushing against it, it is known as a reaction force. The energy for the jump is supplied by the jumper, not by the curb. The jump takes place because the person is pushed by the reaction of the thing he is pushing against, but this reaction is caused by the action provided by the force of the jumper as he prepares to jump.

FRICTION | the force that resists gravity

It is difficult to walk on ice or on a highly polished marble floor. It is not easy for wet hands to pick up a piece of soap. These difficulties result from a lack of friction. Friction is principally responsible for the dissipation of a great deal of energy in the form of heat, and it is as indispensable to everyday living as it is to the functioning of many mechanisms. Traveling is impossible without friction. Because of friction between the wheels of a train and the rails and between the train and the air, a train must consume energy in order to move forward at a uniform speed, rather than doing so on inertia as based on the first principle of dynamics. But it is also true that brakes could not function without friction; even such a drastic measure as jamming steel rods through the wheels would not stop the train. The wheels, even if they stopped turning, would simply slide along the rails. Obviously, then, friction may be either useful or undesirable.

Physics is concerned with, among other things, the general mechanical laws of friction; these laws will be examined

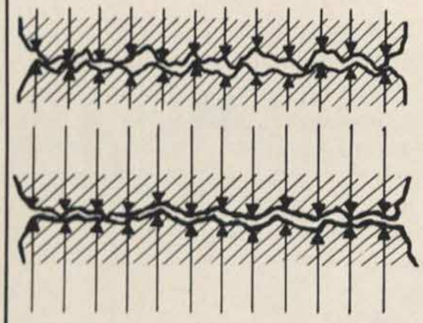
1



THE DEFINITION OF FRICTION—Before discussing friction, it is necessary to understand the meaning of the term. Illustration 1a shows a smooth body (smooth means not excessively rough) resting on a plane that can be inclined at various angles. Here the plane is only slightly inclined, and the body on the plane remains stationary because of friction. When the angle of the plane is increased, the body begins to slide down the plane (Illustration 1b). Holding the angle of inclination at this point halts the body and forms the triangle **ABC** as shown. The ratio between **CA** and **CB**—that is, the ratio between the vertical and the horizontal sides of the triangle—measures the coefficient of friction. No body can remain at rest on an inclined plane if the inclination is more than 45° with respect to the horizontal (Illustration 1c). If a body does succeed in remaining at rest at greater angles, the force that keeps it in this position is not friction. Friction, therefore, is a force that opposes gravity (also, in other cases, it can oppose other forces) and acts in a sense parallel to the plane on which the body is resting. It is therefore a passive force; that is, it does not set bodies in motion but serves only to carry out the work of resistance—the work that opposes motion of bodies. The coefficient of friction measures the maximum slope of the plane on which a body may remain at rest by virtue of friction alone and can never exceed a value of 1 (which corresponds to a slope of 45°) without the intervention of other forces to cause adhesion. A simple example of such a force is glue.

2

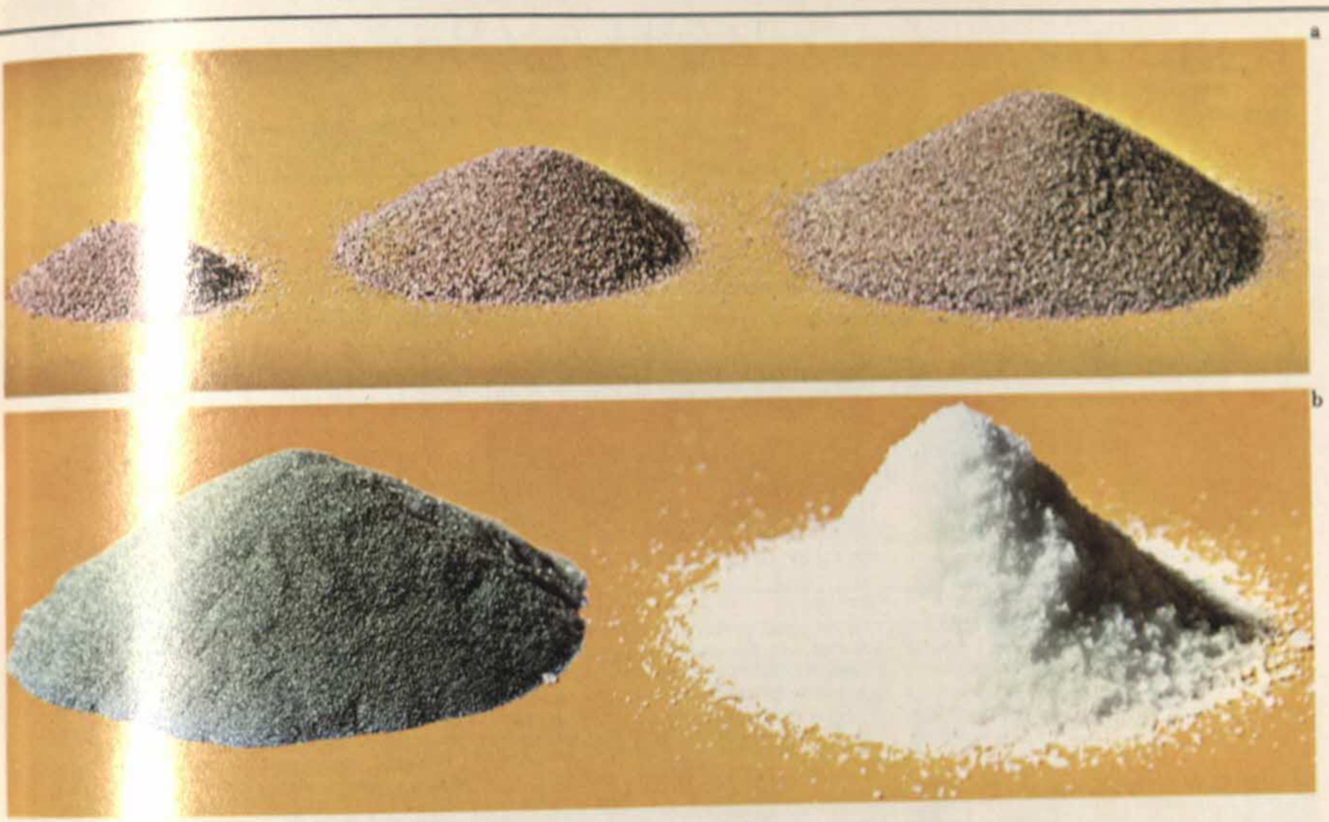
THE CAUSES OF FRICTION—The causes of friction are clarified by this illustration. The surfaces of the bodies in contact are rough; the roughness is irregular, preventing the bodies, even when they may appear to be smooth, from making contact over the whole of their surfaces. Contact is, therefore, made between limited areas, and the weights of the objects are concentrated in these areas. If the pressure is increased between the bodies in contact, they will undergo a slight elastic deformation and touch each other at a larger number of points, thus exerting a greater force on each other.



Increasing contact pressure thus increases the friction. However, no increase in the coefficient of friction occurs. This coefficient is defined as the ratio between two quantities (friction and pressure), which remains constant even if the two quantities vary.

FRICTION AND SURFACE AREA—These three blocks of wood all have the same volume and density, and so have the same weight. However, they constitute three parallelepipeds with different base areas; yet the same force of friction has to be overcome if they are to be dragged. This fact introduces an important law of physics: friction is independent of the surface area of bodies having identical compositions and weights. Friction is always the same for a given weight; therefore, the friction doubles if the weight of a body is doubled. Thus, friction is proportional to the force that two surfaces exert on each other and is independent of the area of the surfaces in contact.

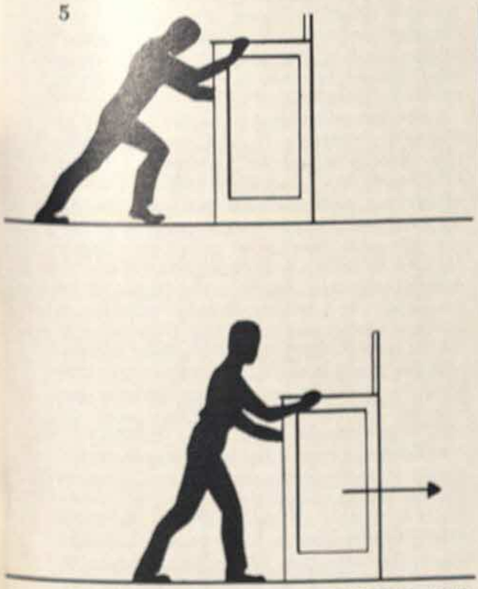
3



4

THE FRICTION CONE—The properties of adhesion of bodies of microscopic dimensions are demonstrated in the properties they display when they are in powder form. In constructing a cone with a small quantity of fine, dry, homogeneous powder (that is, powder formed of equally sized particles) the cone will always have the same slope. The three friction cones shown in illustration 4a are made with the same substance (an abrasive) and, for this reason, always have the same slope. The smaller the friction between the particles of a particular substance, the flatter the cone that can be made from it. Illustration 4b shows two friction cones; the substance on the right (kitchen salt) has greater friction than the substance on the left (sand), and, therefore, forms a steeper cone. It is not possible to construct cones having an angle of more than 45° by relying on friction alone. Thus, the height of a cone obtained in this manner divided by the diameter of the base measures the coefficient of friction of the material used for constructing the cone.

5



STARTING (OR STATIC) FRICTION AND SLIDING FRICTION—Because of the different adhesions established between surfaces when at rest and when in motion, it is harder to start moving a heavy piece of furniture than it is to keep it in motion. Unless otherwise specified, the examples following refer to sliding friction, the more interesting form in all practical applications.

6

COEFFICIENTS OF STATIC FRICTION

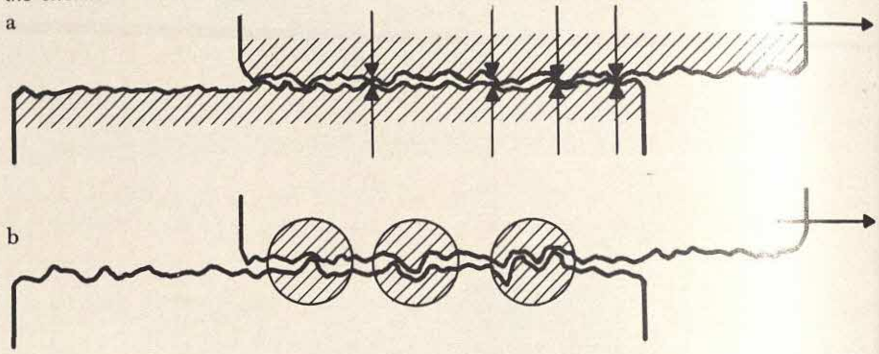
Material	Condition	μm
glass on glass	clean	0.9
glass on glass	lubricated with paraffin oil	0.5
glass on metal	clean	0.6
glass on metal	lubricated	0.2
steel on steel	clean	0.58
steel on steel	lubricated with olive oil	0.105
steel on steel	lubricated with graphite oil	0.13
steel on steel	lubricated with paraffin oil	0.18
steel on bronze	lubricated with mineral oil	0.16
steel on copper	lubricated with mineral oil	0.19
steel on copper	clean	0.35
steel on copper-lead alloy	clean	0.22
ice on ice	clean, temperature 0° C	0.07
ice on ice	clean, temperature −12° C	0.3
ice on ice	clean, temperature −71° C	0.5
ice on ice	clean, temperature −82° C	0.5
ice on ice	clean, temperature −110° C	0.5

THE INFLUENCE OF THE CONTACT-SURFACE MATERIAL—If a block of cast iron is placed on a cast-iron plane or a block of wood is placed on a plane made from the same wood, two different starting angles are obtained (that is, angles wherein the friction fails enough to let the bodies slide). This fact depends on the chemical composition of the materials, which, as will be seen, influences their properties of adhesion. This table gives the values of the coefficient relating to various combinations of seven common substances in different working conditions. Data of this type can be readily obtained from appropriate manuals of physics, chemistry, and engineering.

here. An explanation of the existence of this particular phenomenon will be found, and it will be observed how technology intervenes to increase or diminish friction according to need. In a machine civilization, lubrication is vital because friction wastes energy; an estimated 30 percent or more of the energy generated is consumed in friction.

7

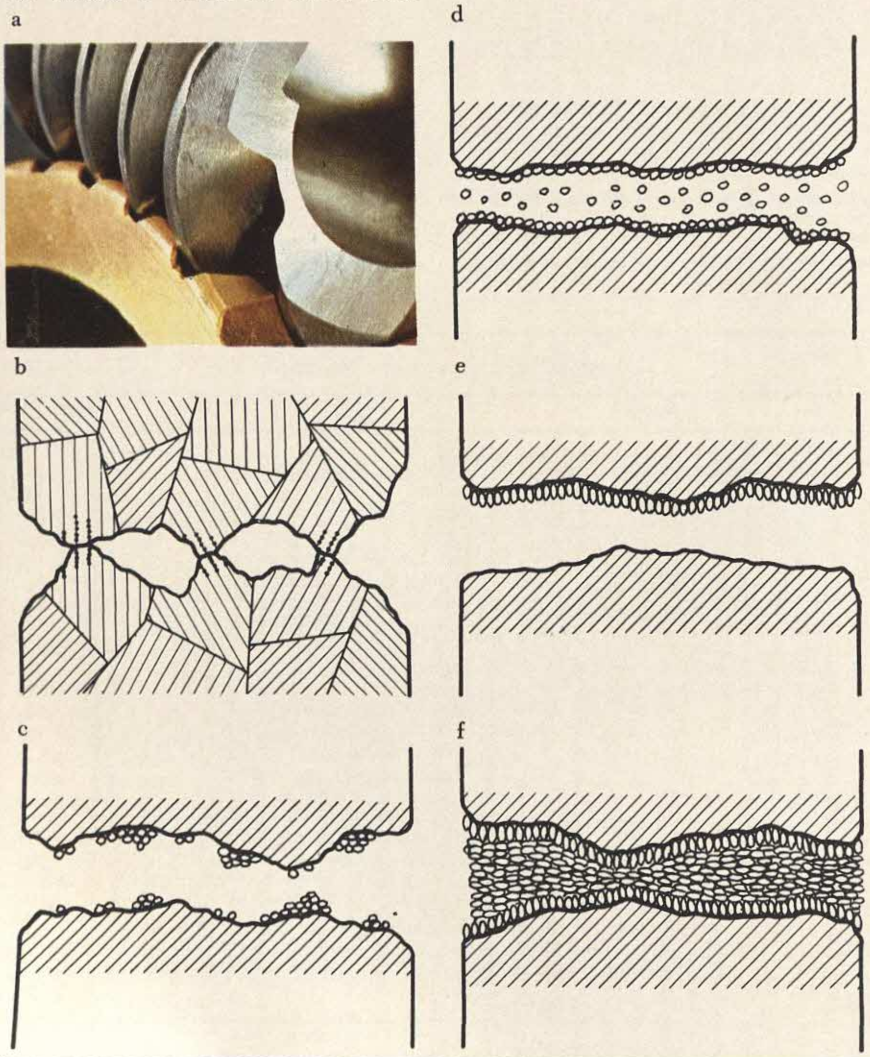
THE ATOMIC ORIGIN OF FRICTION—Illustration 7a shows the cross section of the contact surfaces of two bodies sliding over each other. They are in contact only at a few points, and the pressure is very great at these points. The bodies adhere at these points; that is, their atoms become attached to each other, although only momentarily. The atoms that have become attached to each other must be detached before the motion can continue; this work of detaching the atoms measures the work consumed in friction. Another (microscopic) reason for friction is found in the insertion of protrusions of one surface into the hollows of the other (Illustration 7b); these protrusions must become compressed elastically if the two surfaces are to continue their sliding action. Friction, therefore, depends on the chemical nature of the substances as well as the degree of roughness of the two surfaces.



8

PRACTICAL APPLICATIONS — The general laws of friction have a large number of practical applications.

In mechanisms that involve sliding actions the surfaces in contact are almost always made of different materials, because the adhesion between two such surfaces is less than it would be between surfaces of the same material. In gear teeth, for example, frequent use is made of steel on cast iron, bronze on steel, and so forth (Illustration 8a).



The adhesion between surfaces made of the same material is greater because the atoms at the points of contact can become attached to each other (Illustration 8b).

Friction diminishes in the case of substances whose surfaces are contaminated with some other substance. In fact, this extraneous matter becomes interposed between the atoms of the two surfaces and thus prevents them from coming close to each other and eventually becoming attached (Illustration 8c).

Substances that chemically attack at least one of the two contact surfaces diminish the coefficient of friction even more effectively, because they adhere more firmly to the surface that remains at rest than the one that slides over it (Illustration 8d). The atoms of such a substance form a surface that adheres more to the uneven areas and more effectively prevents adhesion. Gases also act in a manner that tends to reduce the coefficient of friction. In fact, the coefficient of friction is much greater between two clean surfaces in an evacuated container than it is in air.

Large molecules that have the property of chemically attacking at least one of the contact surfaces reduce the coefficient of friction even more; they constitute a velvety layer between the contact surfaces and, consequently, permit a ready sliding action (Illustration 8e).

A lubricant, then, is an agent that greatly reduces friction. It is a substance capable of reacting moderately with the surfaces to be lubricated. A lubricant usually consists of elongated molecules that attach themselves to the surface of the solid by one end, thereby covering it with a kind of molecular "fur." The effect is greater if both contact surfaces are attacked by the lubricant. (Oils, based on fatty acids, attack metals rather strongly.)

When a film of extremely viscous liquid becomes established between the two layers of fur, the sliding of the two surfaces now takes place by means of the layers of the liquid sliding with respect to each other (Illustration 8f). In this case it is misleading to speak of sliding friction between the two surfaces. It is more accurate to speak of the internal friction of the lubricant; that is, the difficulty that its molecules encounter in sliding over each other. This phenomenon is also known as viscosity.

THE CONSERVATION OF ENERGY

a balance sheet that always breaks even

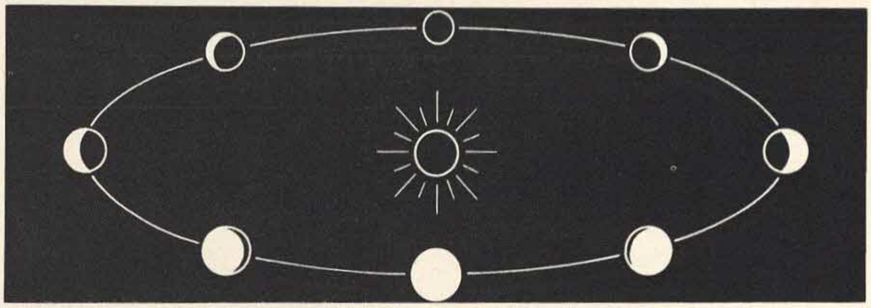
It is easy to believe the presence of energy in a body traveling at high speed. When such a body collides with another, it is apparent that work is done at the expense of the energy of motion—the energy was evidenced by the motion. It is not as easy to perceive the same energy in the case of a stone thrown into the air. Eventually, the stone comes to rest on the ground. The casual observer may see how the energy evidenced by the motion of the stone was partly given up to the molecules of the air that opposed its motion, and finally to whatever it struck when it fell on the ground. It may even seem that the energy disappeared when the stone came to rest, but this is incorrect. Energy is never lost; it is always conserved. This is true of all phenomena that occur in the universe, from atoms to galaxies; the energy present neither grows nor diminishes, but remains constant. All the energy that now exists will never be destroyed—or disappear—but will always remain.

A clear understanding of this universal conservation of energy must begin with the consideration of three concepts: isolated systems; the different forms of energy; and the transformation of energy.

Isolated systems.—A system is considered isolated when it cannot exchange energy with other systems. In actuality, no such system can exist, because anywhere in the universe an exchange of energy with some other system will eventually occur. There are, however, systems that exchange very little energy with the rest of the universe. The planet Venus is an example of such an almost isolated system.

Venus has kinetic energy (or, energy of motion) as a result of its rapid revolution around the sun and its rotation about its own axis. It also has potential energy. Venus is situated at a certain distance from the sun, between the Earth's orbit and that of Mercury. If Venus were to fall into the sun, it would liberate the potential energy it possesses into the gravitational field of the sun, and its potential energy would be transformed into kinetic energy. Moreover, Venus possesses energy in the form of its heat.

1



AN EXAMPLE OF AN ALMOST ISOLATED SYSTEM—Venus represents a system that can be considered isolated. Its energy may change form, but it remains constant and is not exchanged with other systems except to an extremely minor degree.

The various forms of energy possessed by Venus vary little in the course of time—so little that its energy can be considered almost totally constant. The attraction of other planets on Venus is quite small and does not appear to modify its kinetic energy. Its orbit is almost perfectly circular; its distance from the sun is always the same; its potential energy, therefore, remains constant. The energy Venus receives from the sun is radiated back into space sufficiently for the planet to maintain a hot but steady temperature.

Although Venus is not truly isolated, it is sufficiently undisturbed by the remainder of the universe to be a close example of the idealized concept of an isolated system—one that exchanges no energy with the rest of the universe.

Different forms of energy.—Energy can manifest itself in many forms, from the smallest subatomic particles to island universes. For convenience of discussion, energy may be primarily regarded in two forms: kinetic and potential. The energy of planets and solar systems can easily be classified into one of these two forms; it is much more difficult to classify the energy of microscopic bodies.

THE TWO FORMS OF ENERGY IN WATER—The water contained in the water tower is said to have potential energy (Illustration 2a); the powerful jet of water that is issuing from the spout (Illustration 2b) is said to have kinetic energy.

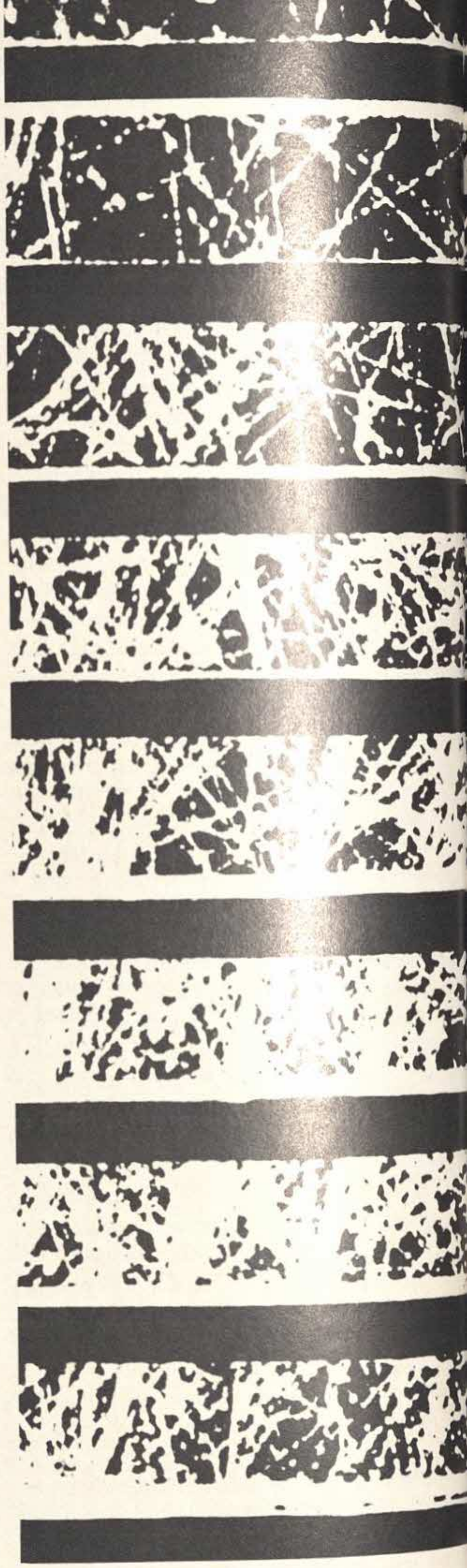
2a

2b

Potential energy is exemplified by the water contained in a storage tower such as shown in Illustration 2. Kinetic energy is exemplified by that same water gushing forth from a pipe or faucet.

THE ELECTROMAGNETIC CASCADE—The light tracks in Illustration 4 (at right) represent the trajectories of electrons and positrons recorded in a cloud chamber with lead baffle plates. When these two particles collide, they annihilate each other and give rise to a gamma ray. Matter has thus been transformed into energy.

4



3a

3b

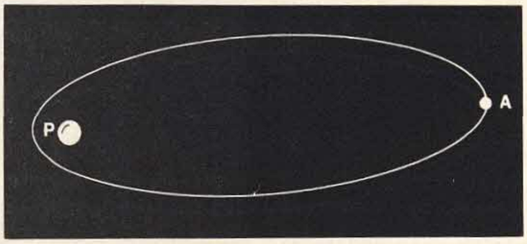
LIGHT ENERGY—A ray of light transports energy. This is demonstrated by using a lens to converge the rays of the sun. In Illustration 3a these rays are concentrated on some sulfur powder. In Illustration 3b the same powder has reached a temperature at which it bursts into flame.

The energy present in microscopic forms of matter may be considered similarly: as potential or kinetic energy. The names for these energies, however, change according to the systems in which they are found. Consider the term *chemical potential.* An example of this is the energy derived from combining potassium chlorate and sulfur powder. These are inert materials; but, if they are mixed and ignited with a flame they combine and produce intense heat. The heat is a form of energy. It was not created by the ignition of the mixed chemicals; rather, it existed in the atoms of the sulfur and the oxygen contained in the potassium chlorate. An attraction exists among these atoms; when they come into close contact, the attraction increases and they become bound to each other. The union of these atoms leads to the formation of molecules in rapid motion—a phenomenon observed as heat. What is termed chemical potential in the world as it is seen by men is potential energy in the microscopic world of the atom. What is termed heat in the world of men is kinetic energy at the level of the atom: a hot body of solid matter consists of molecules in rapid vibratory motion around their equilibrium positions; a hot liquid (or gas) consists of molecules in rapid translatory motion. This microscopic energy is only visible through its secondary effects, but is actually a form of kinetic energy.

Light represents another form by which energy is manifested. A ray of light transports energy. This can be readily observed by the use of a lens to concentrate the rays of the sun (Illustration 3). With a sufficiently powerful lens, substances can be melted or caused to flame.

The transformation of energy.—A phenomenon known as electromagnetic cascade consists of a chain of reactions between positive and negative electrons and gamma rays. When the positive electrons (positrons) collide with the negatively charged electrons of normal mat-

5



THE CONSERVATION OF MECHANICAL ENERGY—The smaller body **A** is revolving around a larger body **P** in an elliptical orbit. As **A** is traveling away from **P**, its speed diminishes; its kinetic energy is transformed into potential energy. As the smaller body approaches **P**, this potential energy is again transformed into kinetic energy as its speed increases. This phenomenon is repeated with each revolution, the energy remaining the same but changing its form as the velocity of **A** increases and decreases. Such a system, therefore, continues its motion indefinitely unless affected by some external force.

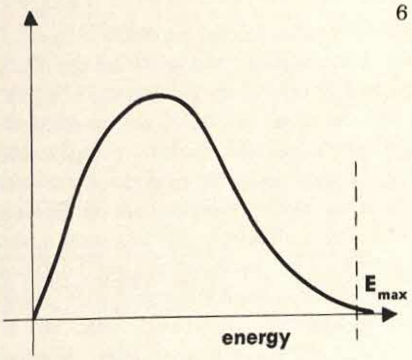
ter, the masses of both particles are totally converted into gamma rays. The energy so released is equal to the sum of the kinetic energies of both the positron and electron involved, plus an amount of energy associated with their mass, equal to the product of their combined mass multiplied by the square of the speed of light. This entire transformation of matter to energy shows that matter itself is a form of energy. The amount of energy released also shows the perfect equivalence of mass and energy, as formulated by the German-born physicist Albert Einstein in his special theory of relativity, which was elaborated in one of his papers that appeared in 1905.

THE CONSERVATION OF MECHANICAL ENERGY

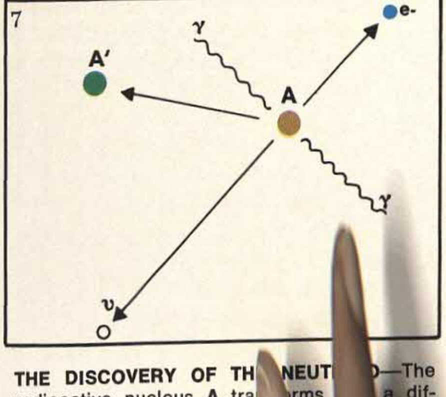
A celestial body gravitating around another in an elliptic orbit (Illustration 7) may appear to behave as if it did not respect the law of conservation of energy. When the smaller mass is moving away from the larger, its speed diminishes as it goes to the outer extremity of its orbit, thus losing kinetic energy. Nevertheless, the total amount of mechanical energy does not change; kinetic energy is merely transformed into potential energy. This potential energy is again transformed back into kinetic energy as the smaller body accelerates on its return movement to its larger companion. This phenomenon continues indefinitely unless some external force changes the motion of the two bodies. If, for example, the same movements occurred within a cloud of gas or dust (such as in a nebula), the atoms of the cloud would become accelerated as they collided with the larger bodies, producing heat. The energy originally represented by the kinetic and potential energies of the orbiting bodies is now transferred in part to the cloud. The transfer represents a loss of energy for the two bodies, but what is involved is only a transformation of energy within the system of cloud and orbiting bodies. The total amount of energy, in accordance with the fundamental law, is conserved.

BETA DECAY—The curve illustrated represents the distribution of energies of electrons emitted by a radioactive nucleus. Although the amount of energy never exceeds a certain maximum value ($\mathbf{E_{max}}$), the electrons are emitted with different energies below this maximum. This phenomenon resulted in discovery of the neutrino (see Illustration 7).

6



7



THE DISCOVERY OF TH[illegible] NEUT[illegible]—The radioactive nucleus **A** tra[illegible]rms [illegible] a different nucleus **A'** by emit[illegible] an [illegible]ron e- and two gamma rays γ. To [illegible]fy th[illegible]iple of the conservation of energ[illegible] was [illegible]ary to suppose that the nuc[illegible] als[illegible] a neutral particle with alm[illegible]ero[illegible] ν. This is the neutrino, nam[illegible]g [illegible]ts existence was experimen[illegible]dem[illegible]

CHEMICAL POTENTIAL—Potassium chlorate and sulfur powder are inert materials when considered individually (Illustration 8a). When mixed and ignited (Illustration 8b), they combine and develop an intense heat. The potential energy previously contained in the atoms manifests itself externally in the form of heat.

8a

8b

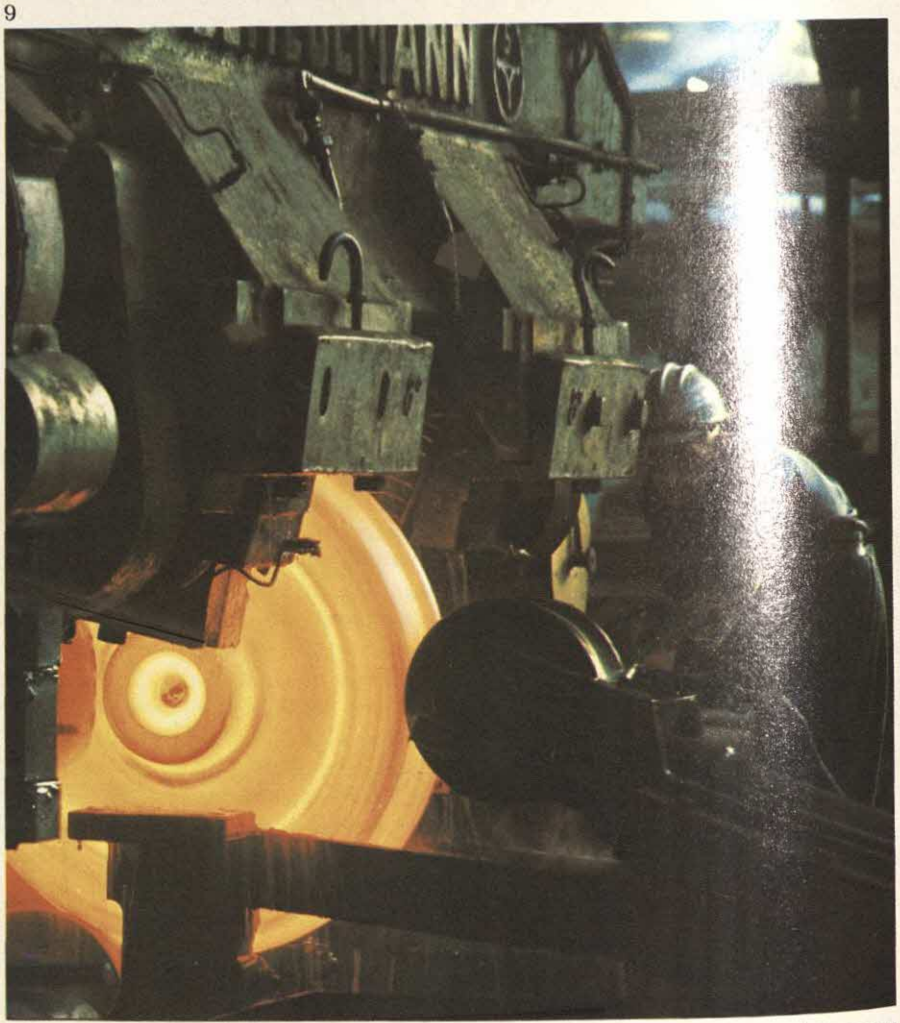
A HEURISTIC APPLICATION OF THE PRINCIPLE OF THE CONSERVATION OF ENERGY

When radioactivity was discovered, it was observed that the nuclei of the atoms of certain elements decomposed by the emission of an electron. This radioactive decay was called beta decay. It appeared logical that identical nuclei should emit electrons having the same amount of energy. Yet, although the energies never exceed a certain maximum, nuclei prone to beta decay emit electrons with different energies.

Gamma rays are also often emitted, and their energies likewise can be measured. No other radiation was observed in these decays. All the nuclei possess the same amount of energy at the beginning; they emit different amounts of energy (an electron and a gamma ray) and come to form new nuclei—again with equal energy content. It was found that the sum of the energies of the final nucleus—the electron and gamma rays—was not equal to the energy of the initial nucleus. This observation contradicts the principle of conservation of energy.

This apparent contradiction was explained by postulating that a third particle—separate from the electron and the gamma ray—was emitted and took away the missing energy. This particle would be extremely difficult to observe, if it in fact existed, because it must be neutral (without electric charge), of extremely small mass (smaller than an electron, whose mass amounts to no more than half a thousandth of the mass of a hydrogen atom), and not likely to react with matter. In the 1930s, the particle was designated the neutrino, and its existence was not demonstrated experimentally until 1960. Nevertheless, scientists were certain of the existence of the neutrino long before its actual observation. The principle of the conservation of energy dictated that it—or its equivalent—exists. This shows that the principle may be used not only to explain known phenomena, but to predict new ones: a heuristic application, after the Greek *eurisco*, meaning "I find."

9



THERMAL ENERGY—A hot body consists of atoms and molecules in rapid vibratory motion around their equilibrium positions. This form of kinetic energy manifests itself through secondary effects; the most obvious of these are heat and incandescence, such as displayed by the glowing ingot in the illustration.

POWER

energy competing against time

The launching of a missile, a huge forging press shaping a piece of incandescent steel, a ship at full speed, a race car in motion, or an airplane breaking the sound barrier, all are manifestations of a mechanical magnitude called power. The common, intuitive concept of power is that of performing a large amount of work in a short time. An athlete lifting a barbell seems powerful because his action and the concentration of his energies are dramatic to the viewer.

It is often said that mankind is searching for large quantities of energy so that its machines may perform large amounts of work. It is more correct to say that man looks for power—for functioning machines capable of liberating large quantities of energy in a short time.

Newer and newer machine tools equipped with more and more powerful engines are constantly being built to perform work at great speed. Numbers of less powerful machines are commonly replaced with one that can produce an equal amount of work. A jumbo jet passenger aircraft does the work of many smaller aircraft; its large engines replace smaller engines. Earth-moving equipment becomes more massive and more powerful as the demand for its use increases.

To understand power, its physical significance must be defined. The definition is necessary because power is a physical magnitude that depends on the energy involved in a given phenomenon and the time in which that energy is expended. The definition of power is, therefore, one of compound magnitude. To be able to analyze this magnitude in various phenomena requires a thorough understanding of the mechanisms of the phenomena. Power will be defined first, followed by an analysis of the power that exists in various phenomena.

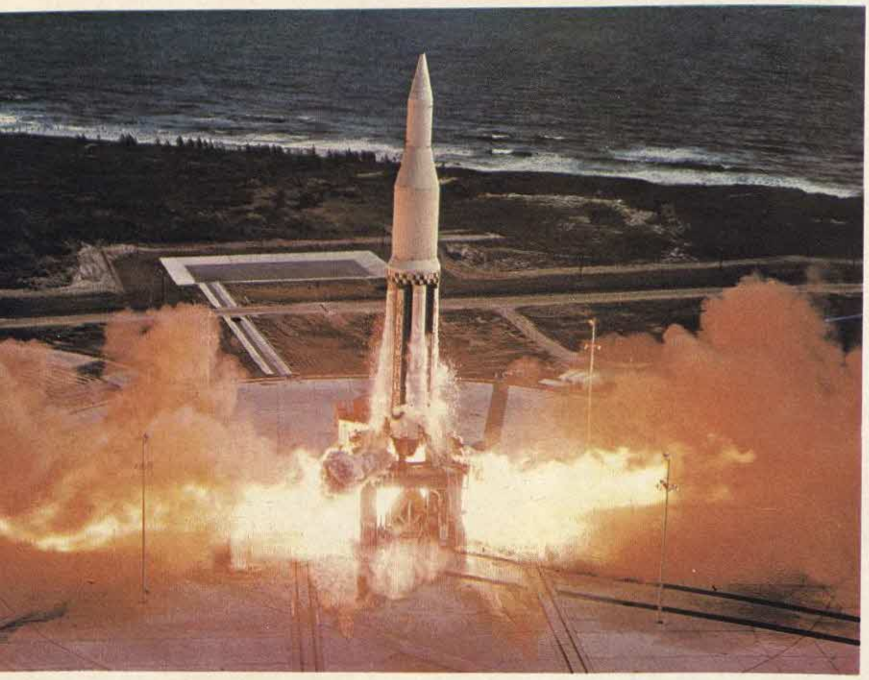
THE DEFINITION OF POWER

Power is the name given to the quantity of energy that is developed in a system in unit time. For example, if a person lifts a weight of 10 kg (about 22 lb) a distance of 1 m (about 3 ft) in one second, the muscles have performed the work of 10 kg-m (kilogram-meters), or about 72 ft-lb (foot-pounds). This work is performed in one second; therefore, a power of 10 kg-m/sec (about 72 ft-lb/sec) is being developed. It is easy to determine the power developed in this case because an exact time for the work is given. If the time employed is greater or less, the calculations are more complicated but still not difficult. It is sufficient to divide the total work performed by the time it takes to perform it. For example, a mountaineer weighing 70 kg (about 154 lb) who climbs 100 m (about 330 ft) in 15 min performs an amount of work equal to the product of the weight of his own body multiplied by the height he has climbed (70 kg × 100 m = 7,000 kg-m); this work is performed in 900 sec. The power that comes into play, therefore, amounts to 7,900 kg-m/sec (about 57,120 ft-lb/sec).

This rigorous definition of power effectively coincides with the intuitive concept that people have of this mechanical magnitude. A shot-putter hurls a heavy ball of iron over a distance of several yards; his muscles contract rapidly, imparting a considerable velocity to the

A CLIMBER—A mountaineer who climbs 100 m (about 330 ft) in 15 minutes performs work amounting to about 7,000 kg-m and develops energy at the rate of about 0.1 hp.

A HIGH JUMPER—In the brief space of time that is needed to make the jump, the muscles of the athlete generate several horsepower.



A ROCKET—The power of the engine of a large rocket, supplying thrust of 100 metric tons (about 220,400 lb) at a speed of 5 km/sec (about 3 mi/sec) is given by the product of the thrust and the speed: 7,000,000 hp.

THE BIRTH OF A MOUNTAIN—The power needed to lift a mountain having a volume of about 1,000 km³ through a height of 1,000 m (about 3,280 ft) in 10 million years amounts to no more than 113 hp.

weight. If he wishes to make another put, he may pick up the weight and carry it back to the point from which he threw it. He is now performing exactly the same amount of work as he did when he threw the weight. The difference lies in the fact that the power in play is great when he throws the weight (because the work is then performed in a short period of time), but is very small when he carries the weight back (because the work is then performed in a much longer time).

HOW POWER IS MEASURED

Many different units of measurement are used to measure power. These units are all obtained by the combination of various units of energy with a unit of time, the second. An important unit of power is the watt (W), named after James Watt, the Scottish engineer who developed the steam engine. The watt corresponds to an expenditure of energy at the rate of one joule per second (about 0.737 ft-lb/sec). A man performing normal gestures with his arm and chest develops power of a few watts. In many mechanisms, however, much larger powers come into play and a greater unit of measurement—the kilowatt—is required. The kilowatt (kw) is a unit of 1,000 W. A unit used for evaluation of power in many practical cases is the kilogram-meter/second (kg-m/sec), which amounts to about 10 W. In actual practice a multiple of this unit—more precisely, 75 kg-m/sec (746 W)—is used. This unit is called one horsepower (hp). A man can develop 1 hp for several seconds when, for example, he is running up a slope. Greater horsepower can be developed by a man who is jumping: the power is developed in the brief instant in which the man contracts his muscles and pushes against the ground as he rises. An exceptionally good high jumper can raise his center of gravity some distance above the position it occupies while he is running up to the bar. This occurs as the result of the impetus he gathers in a fraction of a second; at that instant, the muscles of the jumper develop several units of horsepower.

ANOTHER WAY OF EVALUATING POWER

Assume that a car is being pushed with a force of 10 kg (about 22 lb) to set it in motion. If the same thrust is desired after the car is moving, then increasingly greater power has to be used. In fact, the power is the amount of work performed in unit time. It becomes obvious that, as the car gathers speed, the force applied to it gradually performs more and more work in unit time. When the car is standing still, the applied force does not perform any work; the power is zero. When the car is moving slowly, say, at a speed of 10 cm/sec (about 0.3 ft/sec), the work performed in a second (remembering that work equals force times displacement) is equal to 10 kg × 0.1 m, which is 1 kg-m/sec. However, if the car were to move at a speed of 1 m/sec (about 3 ft/sec), the power would be 10 times greater, because the work performed in 1 second by the thrust is now 10 times greater. Therefore, it is also possible to define power as the product of

AN AUTOMOBILE—The engine of the average automobile develops about 200 hp.

THE POWER OF THE STARS—There are stars that, by virtue of the transformation of hydrogen into helium, supply energy at the rate of as much as 500,000 million billion hp.

the force (which produces movement) and the speed with which this force displaces its point of application. This permits the evaluation of power of a large rocket engine when, for example, it is supplying a thrust of 100 metric tons (about 220,400 lb) at a speed of 5 km/sec (about 3 mi/sec)—a performance typical of rockets used to orbit satellites. The power of such an engine is obtained by multiplying the thrust (100,000 kg) by the velocity (5,000 m/sec). The result is 500,000,000 kg-m/sec. To obtain the result in horsepower, this figure is divided by the number of kg-m in 1 hp (75). The engine used as an example would, therefore, have about 7,000,000 hp—approximately equivalent to the combined horsepower of 30 transatlantic liners.

THE SMALL POWERS INVOLVED IN OROGENESIS

An observer at the foot of a mountain is bound to conclude that only gigantic forces could have lifted such a huge mass to its lofty height. This is correct, but it is wrong to think that large powers were needed. An easy calculation demonstrates this:

First, a mountain having a volume of 1,000 km^3 must be visualized. Assume that it was lifted 1,000 m (about 3,280 ft) over a period of 10 million years. The power per second to raise that mountain to the required height is equivalent to only 113 hp.

Such a mountain would weigh about 3^{15} kg (about 6.6^{14} lb). The work required to raise this mass 1,000 m (about 3,280 ft) would be on the order of 3^{18} kg-m. Since there are about 30,000,000 sec in a year, the time in which this work is performed is about 3^{14} sec. The power is equal to the quotient of the number of kg-m divided by the number of seconds in which they have been expended. The result is 10,000 kg-m/sec. This is equivalent to 133 hp—about as much power as that developed by a small automobile. In fact, the power used to light the cities in the Po Valley of Italy is, in principle, more than sufficient to lift the nearby Alps in only a few hours—something that has been accomplished naturally over a period of several million years of the Earth's history.

THE FORMIDABLE POWER OF THE STARS

All the stars, and this includes the sun, are sources of tremendous energy. They transform hydrogen into helium, a process that occurs at the center of the star. The sun generates energy at the rate of 5^{20} (500,000 million billion) hp. This figure seems fantastic, but there are many stars in the Galaxy that supply energy at a much greater rate—stars that have power outputs as much as ten thousand times greater. Elsewhere in the universe are stars of a different type whose internal energy is liberated at an even greater rate. These stars differ from normal stars but resemble them in certain characteristics. They were, therefore, first referred to as quasi-stellar radio sources, a term that has been subsequently abbreviated to quasars.

IMPACT PHENOMENA

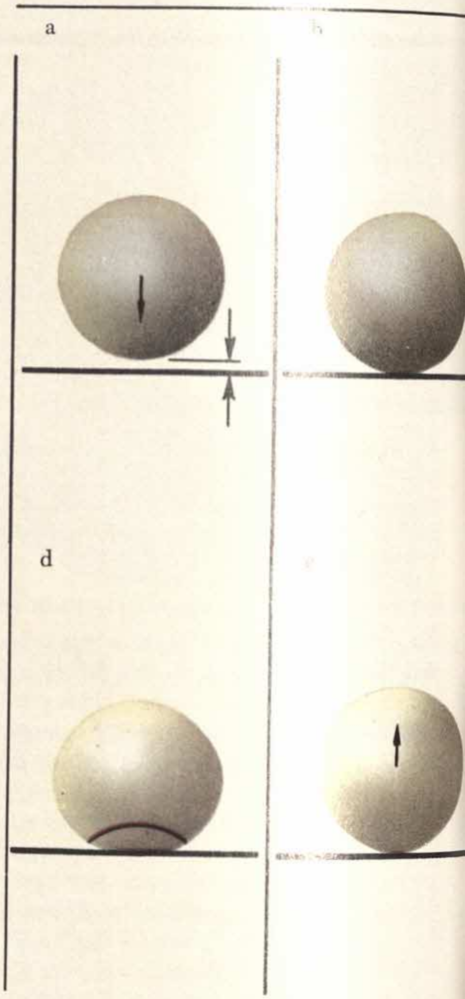
what makes the ball bounce

Many daily experiences involve the phenomenon of impact: a hammer striking the head of a nail, a rubber ball bouncing, two automobiles colliding. In order to understand the meaning of the concept of impact (or collision) in the study of physics, however, it is necessary to analyze impact phenomena in detail, for collisions play a role between the nuclei of atoms and among the stars.

IMPACT ON A SCIENTIFIC LEVEL

When a steel ball is dropped to a pavement from a height, say, of about 20 inches, the ball falls toward the pavement, a sharp sound is heard as it strikes, and the ball rebounds. Although the actual striking of the pavement may be considered the impact, the actual impact from a scientific point of view involves the entirety of the phenomenon, which Illustration 2 shows in slow motion as a sequence of the events once the ball is dropped. In Illustration 2a the ball is about to strike the pavement; it is perfectly spherical in shape. In Illustration 2b the ball is shown at the moment it strikes the pavement, at the maximum velocity it has achieved and just as it is about to lose that velocity because of its collision with the pavement; it is still perfectly spherical. In Illustration 2b′ the part of the ball nearest to the pavement at this point is enlarged 50 million times to show the component atoms. These are spaced at some distance from one another, and are held in position by intermolecular forces represented here by springs. In Illustration 2c the ball is shown at the moment after impact, which causes deformation of the ball at the point of contact. This deformation, which involves compression of part of the sphere, consumes some of the energy of motion (kinetic energy) that the ball had up to this moment. In Illustration 2c′ the same section of the ball, enlarged as before, is shown just after the impact and after the deformation has started. The atoms nearest the point of contact with the pavement are forced back against the atoms further toward the interior of the ball in such a way that compression occurs, as it would if the layers of atoms were linked by springs. As the impact of the ball with the pavement continues, the compression of the "springs" continues, involving more force, and resulting in the conversion of more kinetic energy; the motion of the ball is slowed and ultimately stopped.

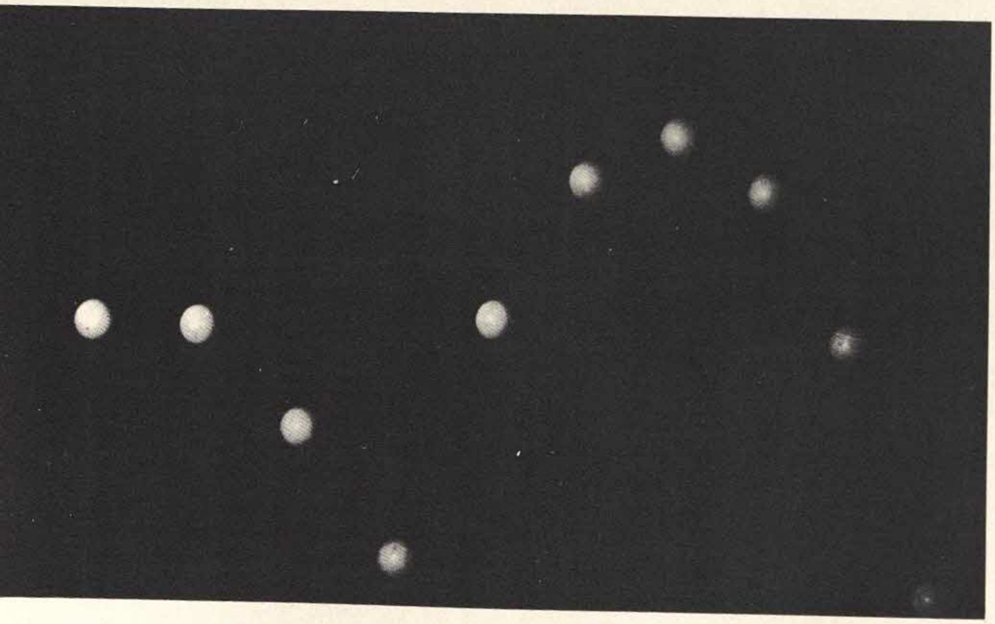
2



ANALYSIS OF IMPACT—In Illustration 2a, the ball is about to touch the pavement; its shape is perfectly spherical. At the instant of impact (Illustration 2b), it is still spherical. In Illustration 2c the part of the ball that touches the pavement is compressed. In Illustration 2d the compression is at its maximum. In Illustration 2e the ball tends to return to its original shape;

THE BOUNCE—A small ball falls to a pavement and bounces. The photo, made with the help of a stroboscope, shows what happens at regular intervals of time as the ball bounces —the movement is from right to left.

1

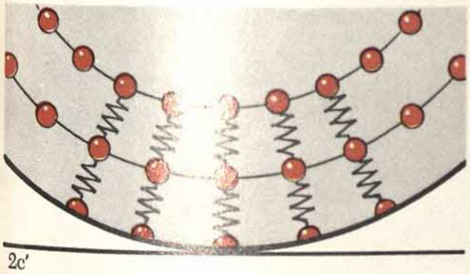


In Illustration 2d the ball is shown at

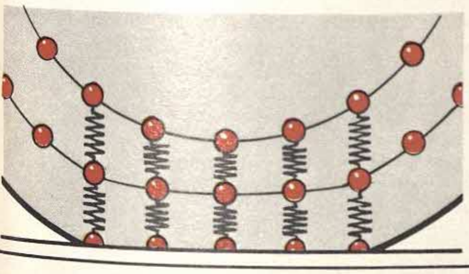
c

The compressed part expands and the ball is forced upward. Illustration 2b′ shows that before the ball has been crushed by the impact, all its atoms are separated from each other by even spaces. Where the ball is compressed (Illustration 2c′), the atoms are forced to come together. Then, the original condition tends to be restored (Illustration 2e′).

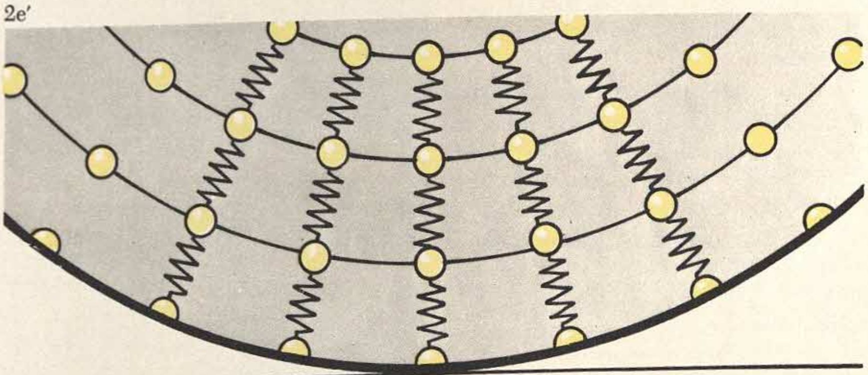
2b′



2c′



2e′



its maximum compression, after expending all its kinetic energy. A steel or glass ball would be compressed much less than the amount indicated in the illustration. A 20-in. fall would result in a deformation on the order of a thousandth of an inch. In Illustration 2e the ball is shown after the impact, about to rebound. After coming to a stop from its fall, the ball resumes its spherical shape because, as shown in Illustration 2e′, the atoms tend to return to the normal distances they had before the impact; the atoms spring apart, and in so doing they exert a force upward on the atoms toward the interior—a force that results in a bounce. Of course, all this has assumed that the pavement was itself not subject to these forces; actually, the pavement undergoes deformation in a manner analogous to that of the ball.

An impact that occurred precisely as indicated in the illustrations would result in a rebound of the ball to the exact level from which it had fallen. In fact, the ball began with a certain amount of potential energy because of its height above the pavement at the beginning of its fall; this potential energy was transformed into kinetic energy (energy of motion) during the fall, theoretically equal in magnitude to the potential energy. On impact, the kinetic energy was transformed into compressional energy while the sphere was deformed, but was restored once the spherical shape resumed; it should, therefore, have carried the ball up to its original height. In reality, the phenomenon develops somewhat differently. Observation of such a situation shows that each rebound is somewhat lower than the previous rebound, indicating that some energy is lost at each impact. What happens to this energy?

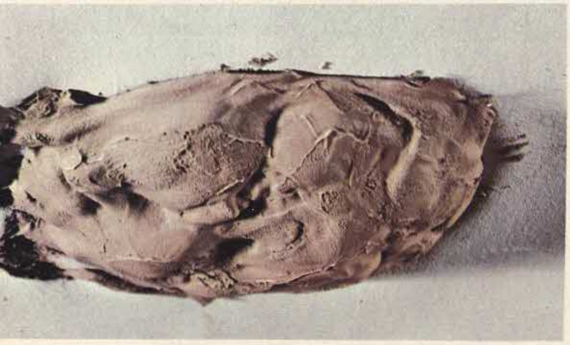
In Illustration 2c′ the deformation of the layers of atoms is illustrated; on impact, the layer nearest the ground drove the second layer upward; that layer drove the third layer upward, and so forth. This occurred suddenly, and not only displaced the layers of atoms in closer proximity to one another, but also made the individual atoms vibrate around their positions of equilibrium. Vibration produces heat; and it is into heat, or thermal energy, that part of the mechanical energy of motion was transformed on impact. This form of energy is not transformed back into kinetic energy, so from the point of view of the ball's bounce, this energy can be described as lost.

ELASTIC AND INELASTIC IMPACT

A perfectly elastic impact, which would involve no loss of mechanical energy and a rebound of the object to the level from which it had fallen, is possible only in theory. In actuality, an object undergoing impact is not governed solely by the disposition of its atoms and must expend part of its energy in impact in the form of heat; it will rebound less each time, and the impact will, therefore, be inelastic—in fact, all impacts that occur are inelastic. However, if the percentage of energy lost is insignificant—amounting to less than 5 percent, as in the case of a steel ball rebounding from a steel surface, or a rubber ball from a pavement—the

A COMPLETELY INELASTIC IMPACT—A sphere of modeling clay, allowed to fall about six feet to the ground, is completely flattened, and does not bounce.

3



impact may be considered relatively (quasi) elastic.

All impacts observed directly are inelastic—that is, those that occur in everyday experience, macroscopically. At a microscopic level, however, an elastic impact may occur. The molecules of a gas, for example, move in straight lines and frequently collide with one another; if they are not moving too fast, their impact may be elastic. Such a molecule may have the same quantity of energy after its collision as it had before. Although a deformation of the molecules occurs just as it does in the case of macroscopic spheres, no energy is absorbed because of the impact.

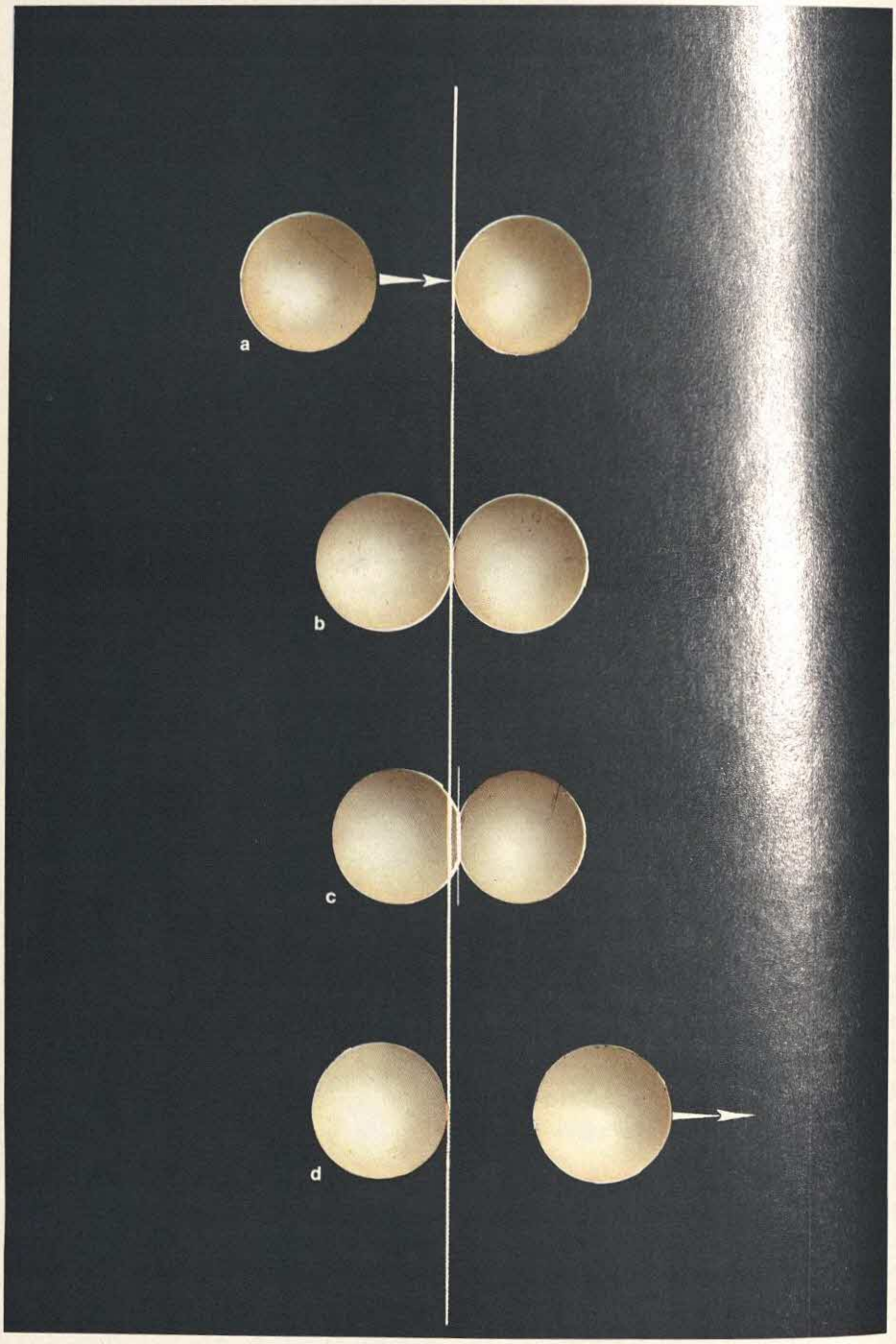
On the other hand, a completely inelastic impact would occur if a lump of modeling clay, for example, falls to the ground from a height of about six feet. Such a lump does not rebound at all; it flattens itself against the ground; all its energy of motion is absorbed in the deformation. Such an impact is almost perfectly inelastic.

IMPACT BETWEEN TWO SPHERES

To study a slightly more complicated impact or collision, two spheres can be used —two billiard balls, for example. One rests on the table; the second ball is driven against it, and if the drive is perfectly centered, on impact the second ball remains motionless in place and the ball that is struck continues to move in the same direction as the striking ball was moving originally, and with the same velocity.

The sequence of events is shown in Illustration 4: In Illustration 4a the sphere on the right, which has the same role as the pavement in the previous experiment, is at rest, and the sphere on the left is about to collide with it; in Illustration 4b the spheres collide, and deformation begins; in Illustration 4c the deformation reaches its maximum as the spheres are driven against each other and then begin to separate, because the energy of motion imparted to the deformation tends to move them apart; and in Illustration 4d the sphere that has been struck acquires

4

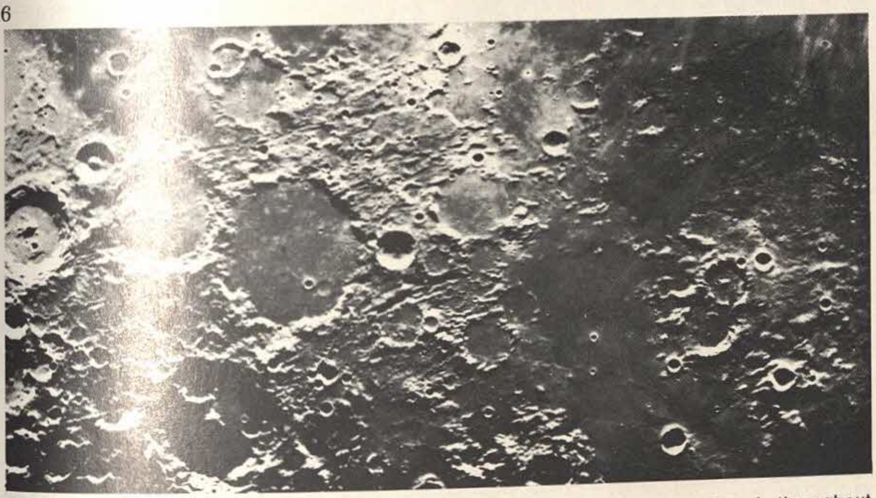


IMPACT BETWEEN TWO SPHERES—If the collision is perfectly centered, the sphere that strikes remains at rest after impact, and the sphere that is struck proceeds on the original course of the striking sphere, in the same direction and with the same velocity.

LENGTH OF IMPACT—The structures of two automobiles that collide while moving at the same speed against each other will be deformed in a matter of hundredths of a second.

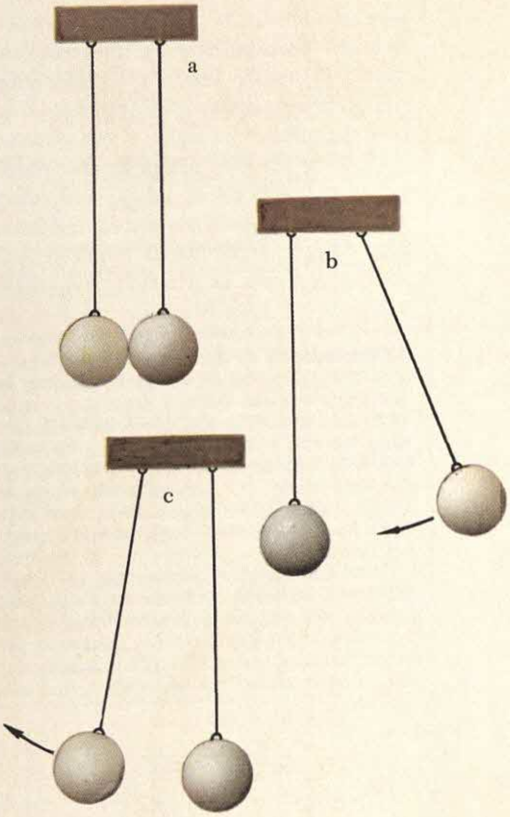
5

6



LUNAR CRATERS—The craters that appear on the surface of the moon result from the impacts of meteorites that travel throughout space at very high speed.

7



ANOTHER IMPACT BETWEEN TWO SPHERES—If two spheres are hanging like pendulums from strings, and one of them is allowed to strike the other, the impact occurs when the string holding the striking sphere has reached a vertical position; the struck sphere will rise to the same height from which the first sphere originally started.

velocity, whereas the sphere on the left that had originally been in motion comes to a dead stop.

In the same way, two identical spheres hanging from a string (Illustration 7a) and free to oscillate like pendulums will behave as shown. If one of them is allowed to drop from a certain height against the other (Illustration 7b), impact occurs when the string holding the first sphere reaches a vertical position. After the impact, the struck sphere rises to the same level, on the opposite side, from which the striking sphere had fallen; and the striking sphere remains motionless (Illustration 7c).

Impact phenomena last for varying lengths of time, depending on the deformability of the objects involved. If two automobiles, for example, each moving at a speed of 95 km/hr (about 60 mph), collide head-on, a time of between one twentieth and one tenth of a second elapses from initial impact to maximum deformation. In the case of a steel sphere striking a steel surface from a height of three feet, only about one ten-thousandth of a second would be required.

COEFFICIENT OF RESTITUTION

In the case of impact between two bodies, one of which is stationary, there is an important relation between the velocity of approach and that of withdrawal. This relation is the coefficient of restitution—the ratio of the velocity of approach and that of withdrawal. The value of the ratio ranges from a minimum of zero to a maximum of one.

HOOKE'S LAW | the measurement of stresses and deformations

The elastic properties of solids, especially metals, are relatively easy to observe; hence they often form an introduction to the study of the properties of matter in general. Subsequently, additional properties—such as internal viscosity and thermal and electric conductivity—can be analyzed. Elasticity and electric conductivity can both be measured with simple instruments. The study of viscosity, however, requires more complex instruments. Acoustic transducers may be needed to measure transit times of penetrating waves. Special attenuation apparatus may be required.

Hooke's law is related to the coefficients of elasticity of solids, and many properties of metals can be deduced by it. Only simple apparatus is required. Beyond the limits of Hooke's law, other properties such as the breaking strength and elongation at failure can be investigated. Properties can be measured by inducing oscillations in a system in which the elastic element that determines them is the metal under examination.

Hooke's law, named for the seventeenth-century English physicist Robert Hooke, states that the stress set up in an elastic body is proportional to the strain to which the body is subjected by the applied load. This law can be used in constructing stress-deformation curves for the study of the structure of a metal.

The proportionality between stress and strain actually does not involve elasticity, which is the property of the solid to recover its original shape completely after the forces producing the deformation are removed. Elasticity, on the other hand, does not necessarily require proportionality between stress and strain. Modern spring steels, the most perfectly elastic materials known, recover their shape completely after deformation but stress and strain do not obey, even approximately, a linear law.

Despite these fundamental conceptual differences between stress-strain proportionality and elasticity, the two concepts are used indiscriminately in such a way that elasticity implies proportionality and vice versa. The theory of elasticity paradoxically is in fact not primarily concerned with elasticity, that is, with the behavior of solids after removal of the load, but is essentially employed as a linear theory with proportionality when the load is applied.

The true significance of the so-called tensile properties is not easily interpreted. Nevertheless, engineers often judge materials through tensile tests. Special dynamometers can be used to measure the huge forces needed to stretch a test piece of large cross-sectional area until it breaks, but the size of a test piece can be scaled down to permit smaller tensile stresses to be measured by simpler apparatus. This type of simplified experiment is shown and described in the accompanying illustrations.

1

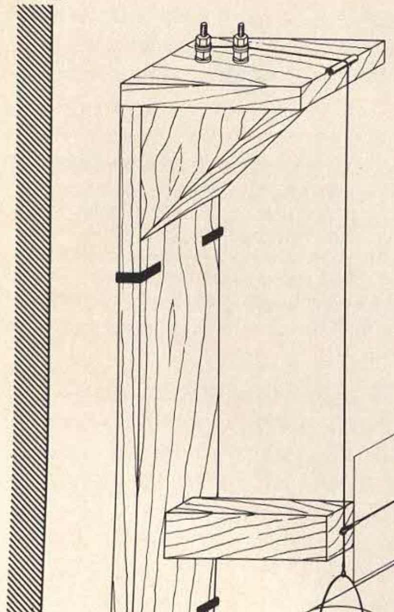
MEASUREMENT OF ELONGATIONS—To measure the elongation of a metallic sample, an apparatus such as the one shown in Illustration 1a is used. The test piece consists of a thin wire with a length of at least 1 m (about 39.4 in.). With strong metals, the diameter of this wire should not exceed a few tenths of a mm (about 0.01 in.); otherwise, very large loads must be imposed to approach the breaking point.

Elongation is the percentage of length stretched. Its value increases with load, but remains relatively small. A wire slightly longer than 1 m (about 40 in.) is most convenient because the most desirable length is exactly 1 m (39.37 in.) between the fasteners at top and bottom.

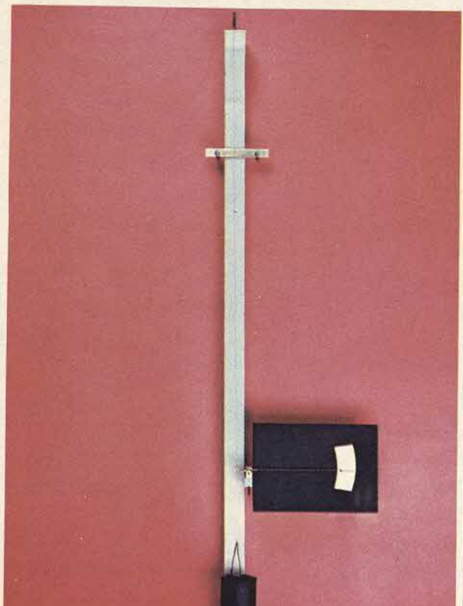
A hardwood board about 150 cm (about 59 in.) long is kept vertical by strong brackets. These brackets are not intended to support weight; their function is to hold the board vertical. On the upper end of the board is a horizontal steel support, over which the wire may slide. The upper end of the specimen is clamped tightly by two bolts.

The specimen is stretched by weights attached to its bottom end. A bucket can be used to hold a large number of weights if the wire being tested is of great strength. Accuracy of measurement is obtained by means of a lever and vernier scale arrangement. An actual test rig is shown in Illustration 1b.

a



b

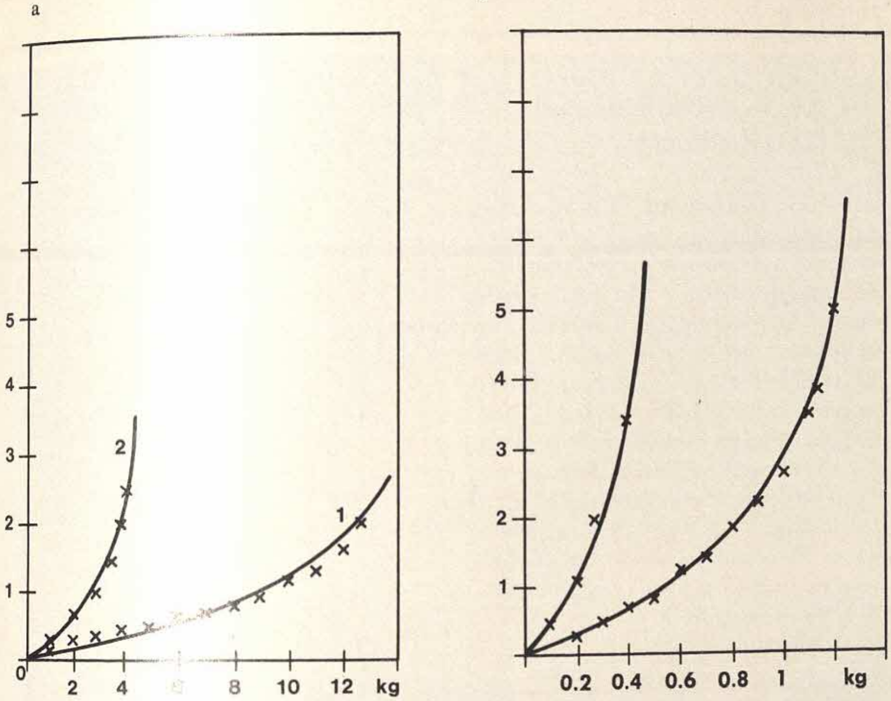


2

HOW TO MAKE THE MEASUREMENTS — The wire used to make measurements must be free of kinks, carefully attached, taut, and carefully stretched. The load is placed at the end of the wire on a plate supported by a wide arc. The wire is initially loaded so that it will be exactly 1 m (39.37 in.) as measured on the vernier scale (as drawn in Illustration 2a and actually installed in Illustration 2b). As additional weight is added, the elongations of the wire are read on the scale. The readings are then plotted against the weights that provide the tensile force.

The maximum weight needed to break the wire should be known. This value is a function of the cross-sectional area of the wire and the nature of the metal. Because steel has a high tensile strength, it must either have a small cross section or be subjected to very large loads in order to break.

3



PLOTTING AND INTERPRETING THE CURVE—The elongations of the wire are plotted as the ordinate and the weights added as the abscissa. Unfortunately, such a curve is not normalized. A better curve would be constructed if the ordinate showed elongation and the abscissa represented the load needed to produce the elongation—with the wire 1 m (39.37 in.) in length and having a cross-sectional area of 1 mm² (about 0.0016 in.²). In such a case, the resulting graph would readily yield Young's modulus: the ratio of the stress to the corresponding strain, used for tension and compression.

The diameter of the wire is, therefore, carefully measured with a micrometer. If a wire has a diameter of 0.3 mm (about 0.01 in.), its cross-sectional area is 0.07 mm² (about 0.0001 in.²). To compare the curve with that of an ideal curve of cross-sectional area of 1 mm², all elongations are multiplied by 0.07.

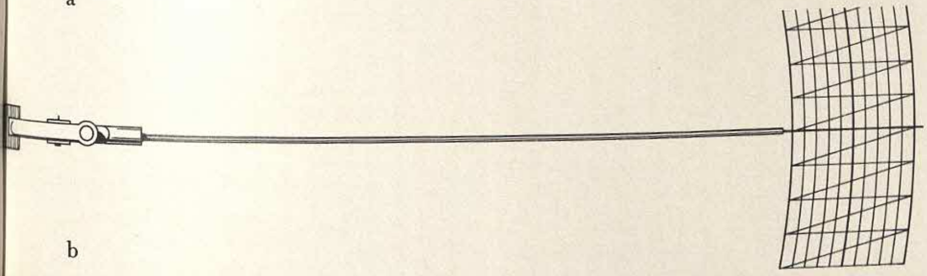
For example, in the straight section of the stress-deformation curve for aluminum (where Hooke's law applies) an elongation of 1 mm is caused by a force of 7 kg (about 15.4 lbs). This is for a wire 1 m long and 1 mm² in cross section. Conversion of these measurements for a wire 1 cm long and 1 cm² in cross section indicates that a weight of 700,000 kg (about 770 tons) is needed to cause an elongation of 1 cm (about 0.39 in.). This is precisely Young's modulus for aluminum.

Repeating tests on different materials can be interesting. The curves shown here illustrate the elastic characteristics of various materials that differ in their constitution, in the elements composing them, or in the heat treatment they have undergone.

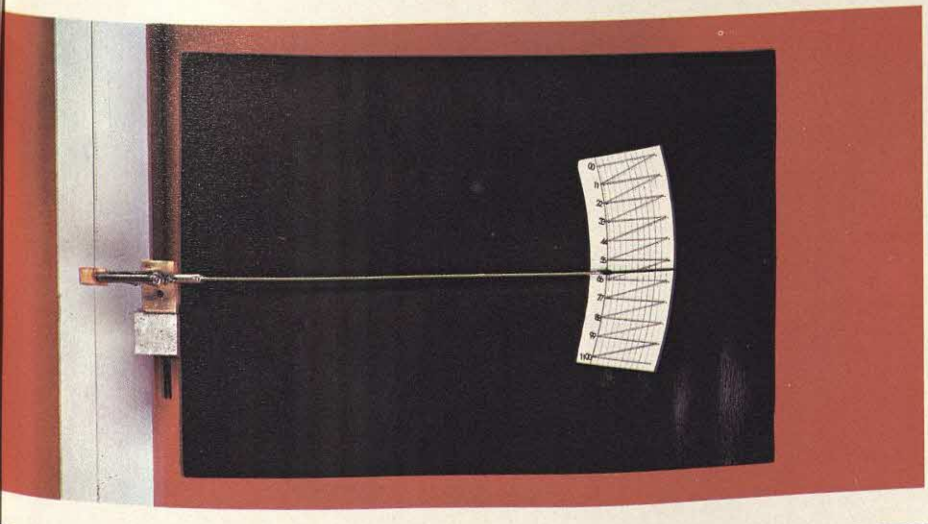
In Illustration 3a, curve 1 represents steel at room temperature; the test was carried out on a wire with a 0.1 mm (about 0.004 in.) diameter. Curve 2 represents steel that was case-hardened prior to the testing. This steel, although it has not undergone any changes in chemical composition, has an enormously smaller breaking load and the part of the curve where stress and deformation are proportional is much shorter.

Metals such as copper, brass, or aluminum display a considerable reduction of tensile strength after annealing.

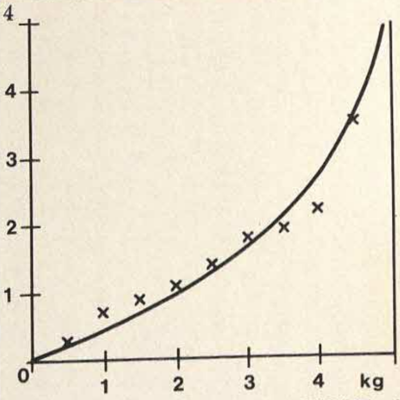
a



b



4



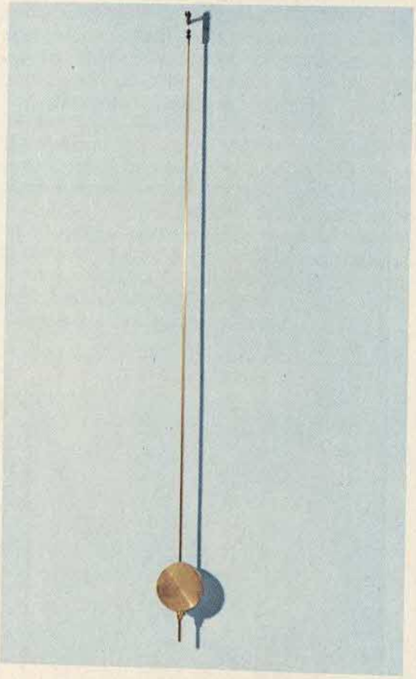
A MORE ADVANCED MEASUREMENT—Another series of measurements is carried out with a wire that is surrounded by refractory material that is heated with an electric resistance. The portion of the wire being heated is sealed off at the ends. The temperature of the wire is determined pyrometrically. Then weights are added and a stress-deformation curve is plotted. The wire is found to be profoundly altered.

THE PENDULUM

simple oscillations and complex figures

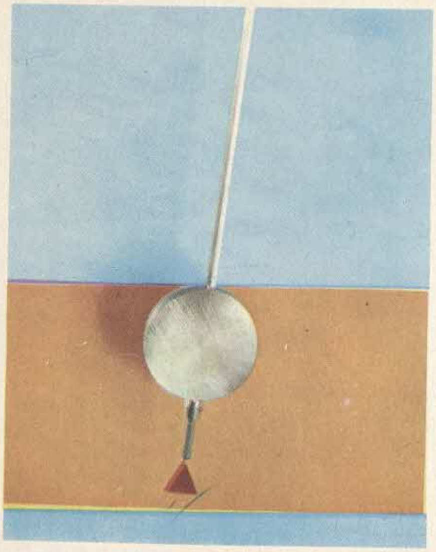
At the beginning of the seventeenth century, in the cathedral of Pisa, Galileo noticed an oil lamp swinging to and fro after a sacristan had set it in motion when lighting it. Galileo watched the oscillations and, using his pulse to measure the time, noted that the time necessary for the lamp to carry out a complete oscillation was constant and was independent of the amplitude of the swing. Later Galileo verified his observation experimentally and suggested that the principle of the pendulum might be applied to the regulation of clocks. Until this event in the cathedral, Galileo had received no instruction in mathematics. After his studies and works in mathematics and science made his name known throughout Italy, he began his research into the theory of motion, first disproving Aristotle's contention that bodies of different weights fall at different speeds. He proved theoretically in about 1604 that falling bodies obey what came to be known as the law of uniformly accelerated motion. The following experiments are designed to demonstrate the properties of a pendulum as formulated in the theory of oscillations based on Galileo's early observations of pendulums.

1



THE PENDULUM TO BE USED IN THE EXPERIMENTS—A simple pendulum consists of a lenticular weight, suspended by a rigid metal rod of small diameter from a thin steel plate fixed to a pendulum support. For these experiments, the support should be, at the least, a sturdy nail anchored in a wall because a flexible support (for example, a wooden rafter that yields as the pendulum moves) would invalidate the results of the experiments. A pendulum about 1 m (about 3 ft) long and whose brass weight is about 1 cm (about 0.4 in.) thick with a diameter of 6 to 8 cm (about 2.4 to 3 in.) should be adequate for these experiments. The steel plate supporting the pendulum should be 5 cm (about 2 in.) long and 2 cm (about 0.8 in.) wide. The thickness of the silicon steel plate must not exceed 0.2 mm (about 0.008 in.). The diameter of the pendulum rod should be approximately 6 mm (about 0.2 in.).

2



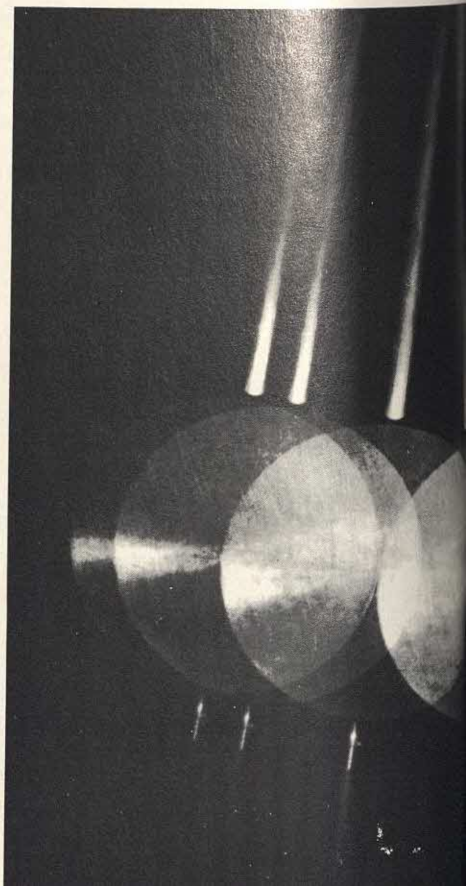
MEASURING THE OSCILLATION PERIOD OF A PENDULUM — In order to measure the oscillation period of a pendulum accurately, the place where the oscillation starts must be clearly marked. In the illustration, the vertex of the triangle marks the point to which the pendulum is pulled and then released. Before the oscillation period can be measured, the regularity of the pendulum motion must be checked to establish that the pendulum oscillates smoothly in a plane. Once the pendulum is oscillating regularly, the period can be measured. A stopwatch with a second hand should be used for accurate measurement. After the regularity of the oscillations has been established, the pendulum should be timed for at least a hundred oscillations so that error in measuring the period, which could be as great as several tenths of a second, is distributed over a considerable number of oscillations and is correspondingly small. Also, the error will tend to grow less as the observer gains familiarity with the procedure and skill in judging the exact time when the pendulum reverses the direction of its movement.

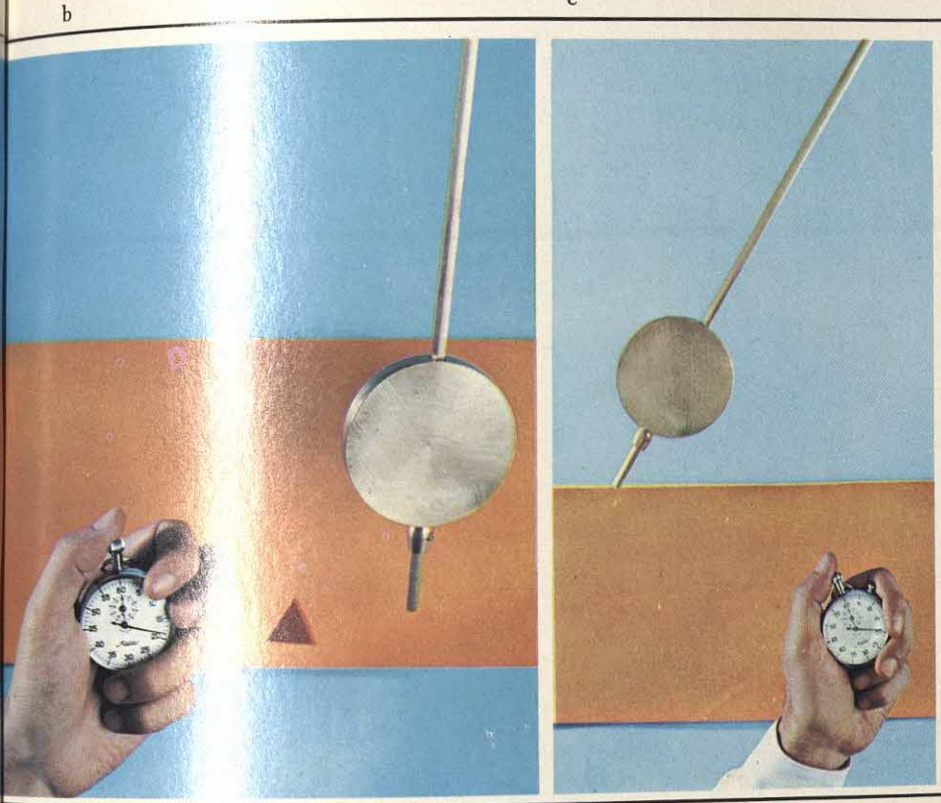
3
a

4

THE SPEED OF THE PENDULUM—The following experiment illustrates the nature of the velocity of the pendulum.

More preparation is required for this experiment than for the previous ones. A source of intermittent light, such as a stroboscope, and a camera synchronized with the stroboscope are set up in a darkened room so that the





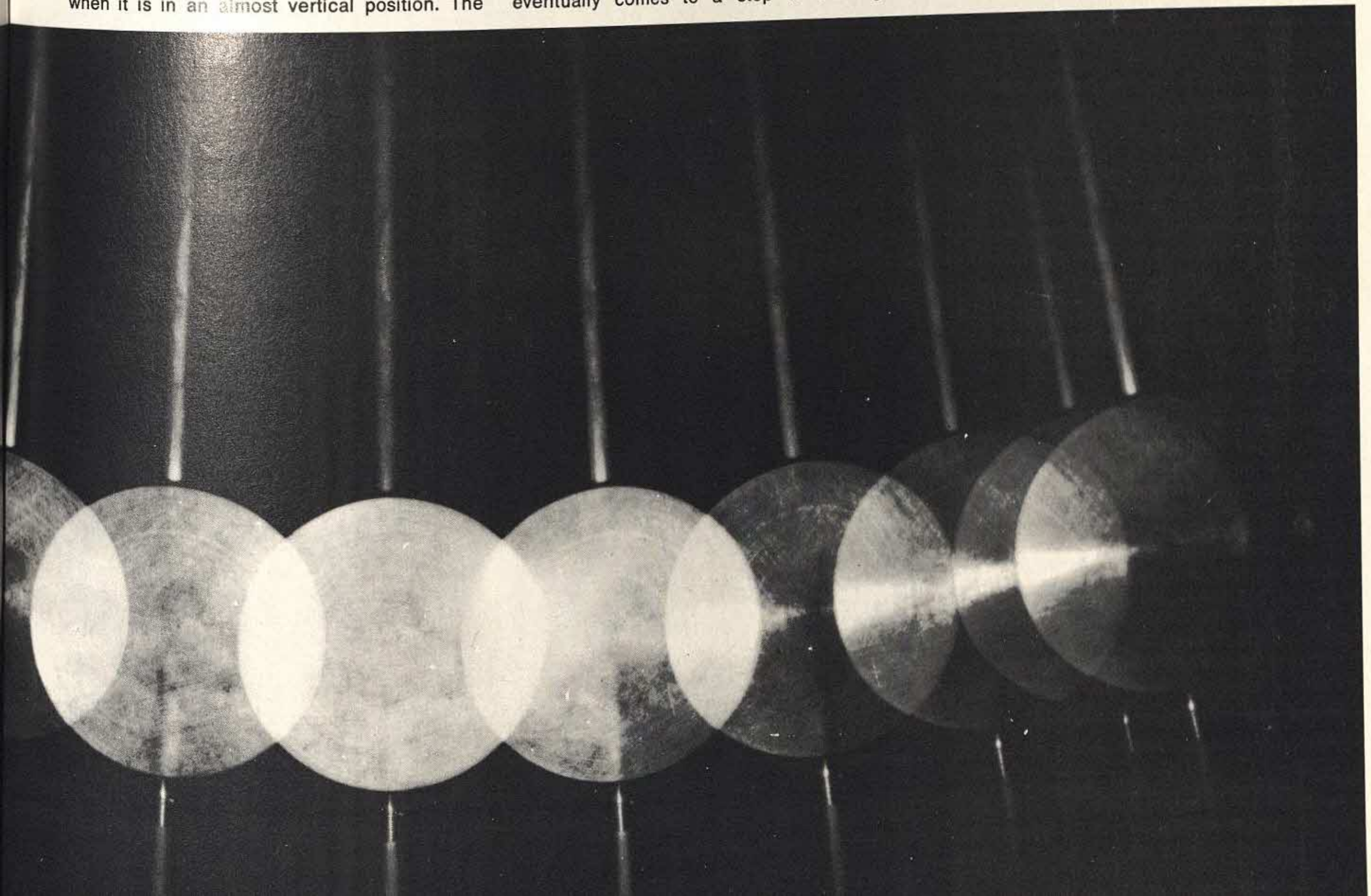
THE ISOCHRONISM OF A PENDULUM—This experiment demonstrates that the oscillations of a pendulum are isochronous; that is, they last the same length of time regardless of the amplitude of the oscillation, if the amplitude of the oscillation does not exceed a certain value.

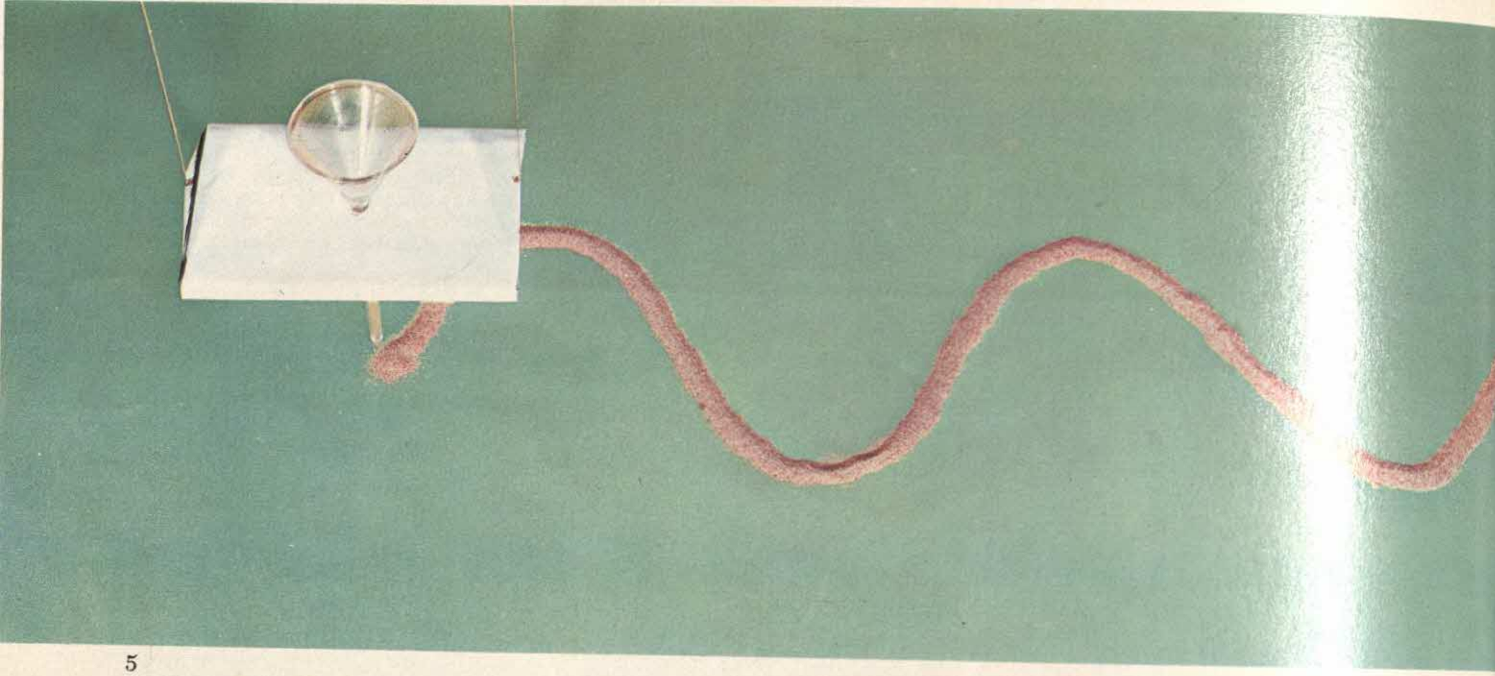
The pendulum is set in motion, as shown in Illustration 3a, so that it will complete at least 100 swings. Then, the time the pendulum takes to complete the first 10 full oscillations is measured. After the oscillations become damped as shown in Illustration 3b, another 10 oscillations are measured. Within the limits of error inherent in the stopwatch, the oscillation times should be equal.

However, if the amplitude of the oscillation exceeds a certain value, the oscillations will not be isochronous. This can be seen in the supplemental experiment shown in Illustration 3c. The pendulum is displaced about 30 or 40° from the vertical, and the time the pendulum takes to complete the first 10 full oscillations is measured. After the oscillations become damped and are about to die out, 10 oscillations are measured. There should be a discrepancy of about 1 second between the two values. This discrepancy is not due to errors in measurement, as the oscillations of a pendulum are no longer isochronous beyond a certain limiting value of the amplitude of the swing.

stroboscope illuminates the pendulum and the camera can take multiple exposures of the pendulum each time the stroboscope flashes on. The resultant photograph will be similar in nature to the illustration. The images of the pendulum bob will be more closely spaced at the ends of the oscillation and wider apart when it is in an almost vertical position. The speed with which the pendulum moves is at a minimum at either end of the swing and at a maximum when the pendulum crosses the vertical.

The pendulum moves at an almost constant speed for a short distance on either side of the vertical, but slows down as it rises and eventually comes to a stop at the highest point of the oscillation. Thus, the acceleration is zero when the pendulum is completely vertical and is at a maximum at the moment when it changes direction at either end of the oscillation. Therefore, the acceleration is zero when the velocity is at a maximum, and the velocity is zero when the acceleration is at a maximum.





5

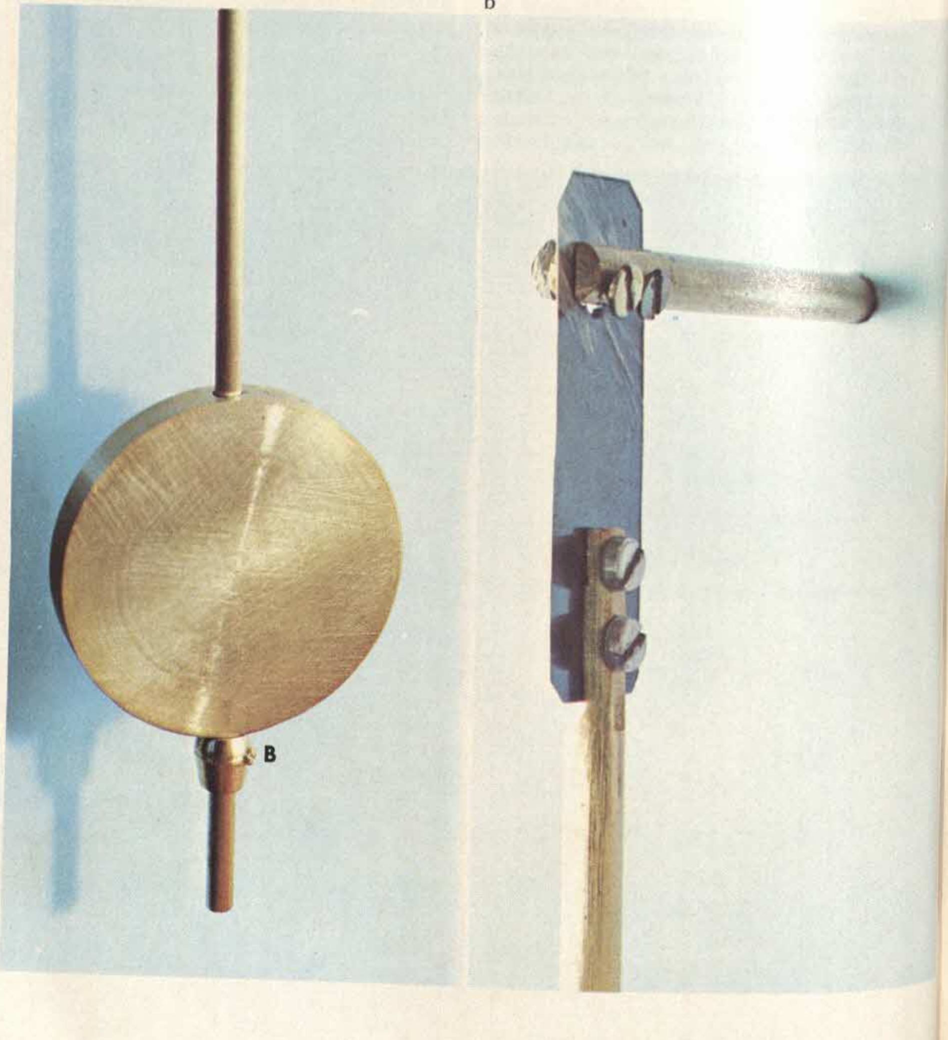
HARMONIC MOTION—This experiment demonstrates the relation between the motion of a pendulum and the graph of a sinusoid.

Using two pieces of string and a small platformlike board, a pendulum that operates on the principle of a swing is constructed. A hole is drilled in the center of the board and a small funnel is inserted in the hole. The funnel is filled with fine dry sand that has been washed with hydrochloric acid to help the sand pour out of the funnel steadily. The swing is then pulled to a low height and released so that it oscillates regularly. The bottom of the funnel is opened to let out the sand. To keep the sand from spilling onto the floor, a long sheet of paper is placed under the pendulum and pulled perpendicular to the plane in which the pendulum is oscillating. If the paper is moved at a uniform speed, the falling sand will form a perfect sinusoidal curve. Such a motion, which traces a sinusoidal curve on a plane moving at a constant speed with respect to the motion, is called a harmonic motion.

6

THE PERIOD OF OSCILLATION AND THE LENGTH OF THE PENDULUM—The pendulum used to study the properties of pendulums is a real pendulum. That is, it consists of a weight with length, width, and height. Its weight does not act through the geometrical center of mass of the bob, as the simplification of the theory usually assumes (Illustration 6a). The pendulum is suspended by a rod with a weight of its own and is attached to a thin steel plate that is not completely flexible (Illustration 6b). Therefore, the results of these experiments will not conform exactly to theoretical predictions of the motion of an idealized pendulum where the mass is assumed concentrated entirely at the bob. However, if care is taken in the construction of the pendulum, the experimental results should be close enough to the predicted results to serve as a means of verifying the theory.

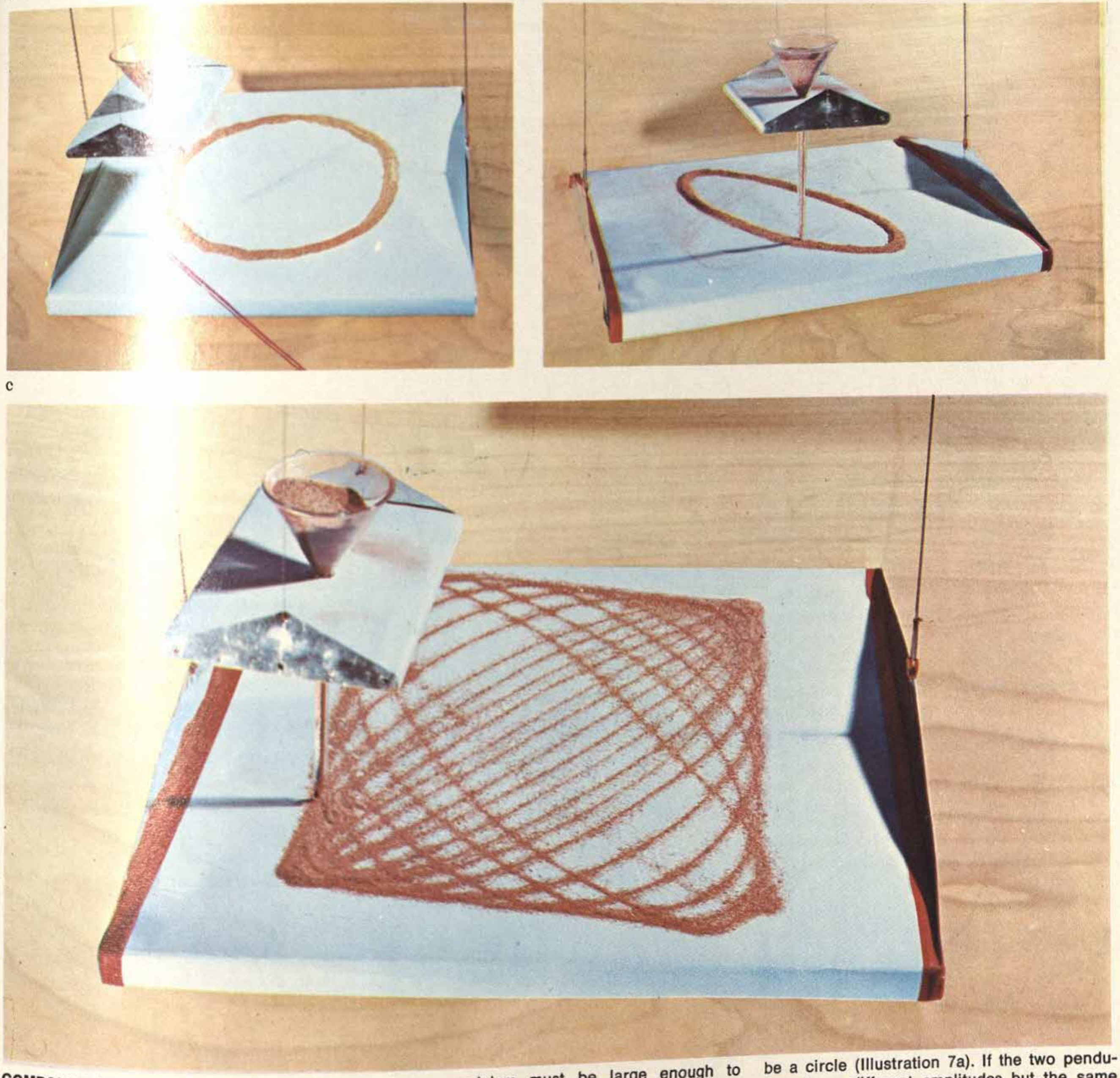
The next experiment verifies the relationship between the length of the pendulum and the period of oscillation, a relationship that is exact for an idealized pendulum, but one that is only approximate for a real pendulum. The relationship is $T = 2\pi\sqrt{(l/g)}$, where T is the period of the pendulum, l the length of the pendulum, and g the acceleration due to gravity. For this experiment, the pendulum should have a weight that can be moved up and down the rod and locked into place. Such a pendulum is shown in Illustration 6a. The locking screw **B** serves to fix the weight at any desired height. If the period is measured in the manner described in Illustration 2 and the length of the pendulum is taken to be approximately the distance between the center of the weight and the point of suspension, the results should show (within the limits of experimental error) that the formula given is correct.



a

b

7



COMPOUND HARMONIC MOTION—Compound harmonic motion can be studied with the aid of two pendulums constructed like the one described in Illustration 5 and set up so that they will oscillate perpendicular to each other. The two pendulums should be mounted as shown above. The upper pendulum should be small enough to oscillate within the space between the strings of the lower one. The upper pendulum carries the funnel with the sand; the lower pendulum must be large enough to gather all the sand falling from the upper one. The periods can be synchronized by attaching weights to the strings of the smaller pendulum and moving them up and down the strings until the pendulum swings with the desired period. If the two pendulums are set in motion, some rather complex figures will result. If the pendulums have exactly the same period and the same amplitude, the resultant figure will be a circle (Illustration 7a). If the two pendulums have different amplitudes but the same oscillation period, the sand will trace an ellipse (Illustration 7b). And if both the periods and the amplitudes of the two pendulums are different, the resulting figure will be of the type shown in Illustration 7c. These complex figures are called Lissajous curves; they are named after Jules Antoine Lissajous, the French mathematician who first studied them.

UNIVERSAL GRAVITATION

a great and mysterious force

In science, as in other areas of life, some things are taken on faith. So it is with the fascinating phenomenon of universal gravitation; scientists understand the "how" of it, but the "why" is still a mystery.

Isaac Newton, who in 1687 published the *Law of Universal Gravitation*, admittedly could find no explanation for the origin of gravitational forces. His famous phrase, "I frame no hypotheses," underscores his concern with a descriptive, rather than an interpretive, approach to the study of gravitational forces.

NEWTON'S LAW

The story of Newton and the falling apple, while possibly apocryphal, dramatically illustrates the importance of Newton's discovery that bodies tend to fall toward the Earth. They fall because they have weight, and the cause of their weight is the attraction that exists between the matter of which they are formed and the matter that comprises the Earth.

This attraction is an all-pervading force that depends only on the masses of the bodies and their distance apart. The Earth, therefore, exerts an attraction on the moon just as it does on the apple. The demonstration that these two particular attractions follow the same mathematical formula was the chief contribution Newton made to the study of celestial bodies.

Newton's law refers to mass instead of weight because the weight of a body on earth is really the particular strength of the Earth's gravitation on that body. On a different planet, the same body would have a different weight, but its mass would remain the same. For example, a body weighs six times as much on Earth as it does on the moon because the moon has a smaller mass than the Earth and therefore creates less gravitation.

Newton's law may be stated as follows: Each particle of matter is attracted to every other particle by a force varying as the product of the masses of the particles and inversely as the square of the distance between them.

Newton's law is formulated as

$$F = G\frac{M_1M_2}{d^2}$$

where F is the force between the two particles of mass M_1 and M_2, d is the distance between them, and G is a constant for a variety of matter and is called the universal gravitational constant. This constant has the dimensions $L^3M^{-1}T^{-2}$, where L is length, M is mass, and T is time, and a numerical value depending on the units used. (If M_1 and M_2 are given in grams and d in centimeters, F will be in dynes if $G = 6.670 \times 10^{-8}$.)

The application of this law to the solar system not only supported all of Kepler's previous laws, but also provided an explanation for the movement of the planets and their satellites—an explanation that was in harmony with observations.

MEASURING THE FORCE OF GRAVITATION

Physicists who study the nature of electromagnetic forces classify the force of gravity among the so-called weak interactions. The following comparisons demonstrate why this is so. If the electrons from the atoms of a handful of matter could be detached from their nuclear orbits, they would be attracted by the nuclei at a distance of a few centimeters by electrostatic forces equal to millions of tons. On the other hand, the gravitational attraction that a handful of matter exerts on another handful of matter at a few centimeters is virtually a negligible force; it could not be measured with ordinary instruments, and only certain extremely delicate experiments could demonstrate its existence, for that matter. Two bowling balls placed one inch apart, for example, would show a mutual gravitational pull of about 0.1 dyne (about 0.0001 g).

The force of gravitational attraction is not to be measured in the laboratory alone, however. In the universe it is the key that unlocks many celestial mysteries and enables scientists to make accurate predictions of the motion of the stars.

APPLICATIONS AND LIMITATIONS OF NEWTON'S LAW

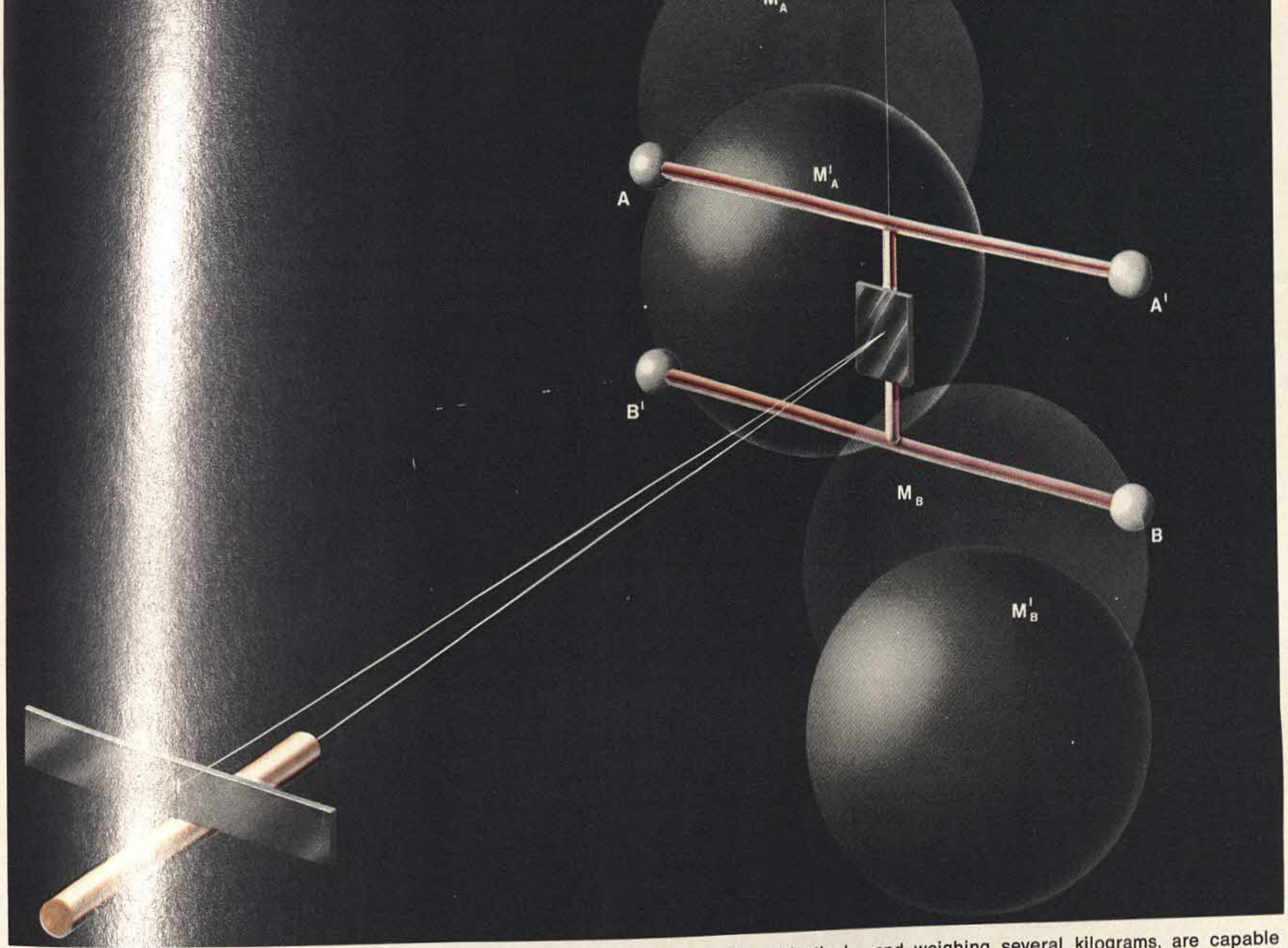
The laws of universal gravitation developed by Newton are the basis on which more complex laws governing a large number of phenomena are built. Some of these phenomena are so complex that they require an exhausting number of calculations in order to be theoretically described. A good example is the motion of the moon.

In order to predict the precise point in the sky in which the moon will be found a year after its position has been accurately measured by experimental means, an enormous number of variables must be considered, including the fact that the lunar orbit is neither perfectly circular nor perfectly elliptical, that the velocity of the moon varies according to its distance from the sun, and so forth. To predict the passage of the moon in a certain point of the sky with an accuracy of 0.1 sec, something like 1,500 correction factors must be applied to the first calculations.

Some measurements are even more complex. To calculate the exact orbit an artificial satellite follows around the Earth, as many correction factors are needed as are used in the moon calculation; and still the orbit will not conform exactly to the one predicted by theoretical calculations. This is due to the fact that scientists do not yet have an exact knowledge of the forces that act on an artificial satellite. For example, the action exerted on an artificial satellite's motion by the continental landmasses is unknown. However, after experimentally measuring the trajectory of the satellite, the deviations can be used for evaluating the mass of the continental landmasses.

Without an understanding of the effect of the forces of universal gravitation, the solution of complex calculations such as those described would not be possible.

In some cases, however, Newton's formula does not seem to be adequate. Such a case is the orbit of the planet Mercury. In analyzing the motion of Mercury around the sun, scientists discovered that the motion of its perihelion



DETERMINATION OF THE GRAVITATIONAL CONSTANT—The determination of the gravitational constant represents one of the most delicate measurements in all physics. It was carried out for the first time in 1797–1798 by Lord Henry Cavendish, an English chemist and physicist who used a torsion balance he had invented for the purpose.

The illustration shows a torsion balance designed by the English physicist Charles V. Boys in the late nineteenth century for accurate measurement of the gravitational constant.

Two small metal spheres **A** and **B,** having a mass of a few grams, are suspended at different heights. Spheres **A′** and **B′** are identical to spheres **A** and **B,** and are used to balance the system. The spheres are initially isolated in space, subject to no appreciable attraction except that of the Earth.

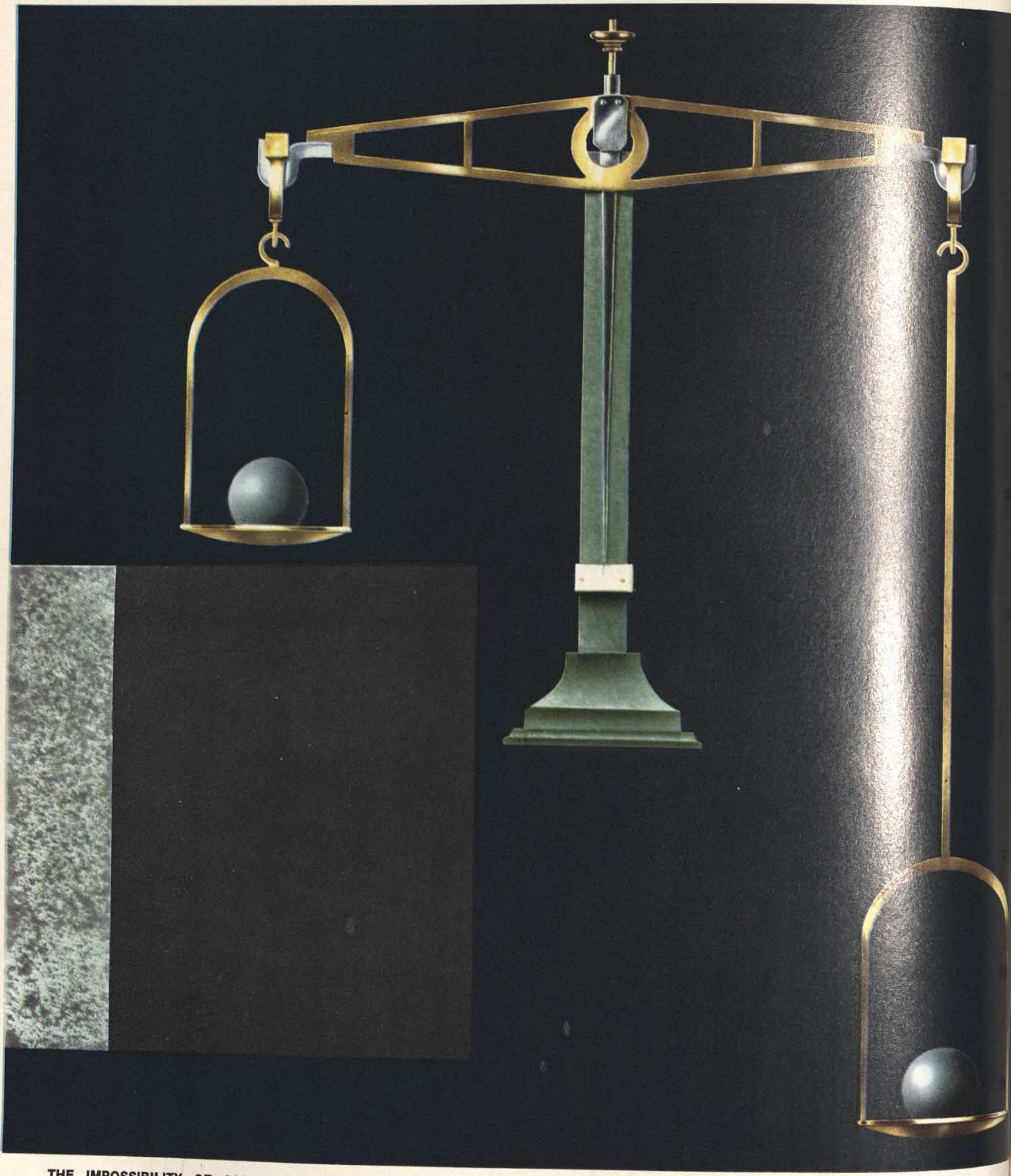
A flat mirror, used to indicate the orientation of the system by means of a method known as the optical lever, is mounted on the yoke. A projector throws a beam of light against the mirror that is focused onto a thin, elongated spot at a specified distance, which marks the zero. Two large spheres, $\mathbf{M'_A}$ and $\mathbf{M'_B}$ are far away. The spheres $\mathbf{M_A}$ and $\mathbf{M_B}$, made of a heavy metal such as lead, and weighing several kilograms, are capable of exerting a gravitational attraction on the small spheres **A** and **B.** The force exerted between the large spheres and the small ones causes the yoke of the torsion balance to rotate. This rotation is opposed by the elastic reaction of the suspension wire. The amount of reaction depends on the amount of rotation and can be calibrated accordingly. The amount of rotation is measured by the displacement of the spot of light on the scale. Once the rotation is known, the force exerted between the large and small spheres can be measured.

(the point in its orbit when it is closest to the sun) showed a greater advance in the direction of the planet's motion than could be accounted for by the gravitational action of all the other planets. This would seem to indicate that Euclidian geometry is not valid in the space very close to the sun—the space in which Mercury moves. This space has a geometry in which the shortest distance between two points is not a segment of a straight line, but rather a curve. Such a deviation from Euclidian geometry would seem to invalidate the universality of the law of universal gravitation. How, then, is this exception explained? The answer came in 1915 when Albert Einstein formulated his general theory of relativity. While some scientists at first thought Einstein's theory disproved Newton's law, it in fact only presented a different description of gravitation. Einstein suggested that objects do not attract each other by exerting a pull. Space, he said, is warped or curved near matter and the curvature is greatest near heavy bodies.

The particular curvature around any body is called the gravitational field of that body. The gravitational field of the sun is a kind of hill in space, which Einstein called space time. The sun is at the top of the hill, and the planets and other bodies that move in the sun's gravitational field go around the sun instead of over it because the curved path is easier to take than the hill.

According to Einsteinian theory, bodies do not attract each other at a distance—they merely follow the line of least resistance through the hills and valleys of the curved space that surrounds other bodies. Objects that fall to Earth are not "pulled" by the Earth; rather they are pushed toward it by the gravitational field.

Einstein, no more than Newton, attempted to explain the cause of gravitation. He, too, refrained from "framing an hypothesis," although in 1929 he suggested that gravitation and electromagnetism might be related.



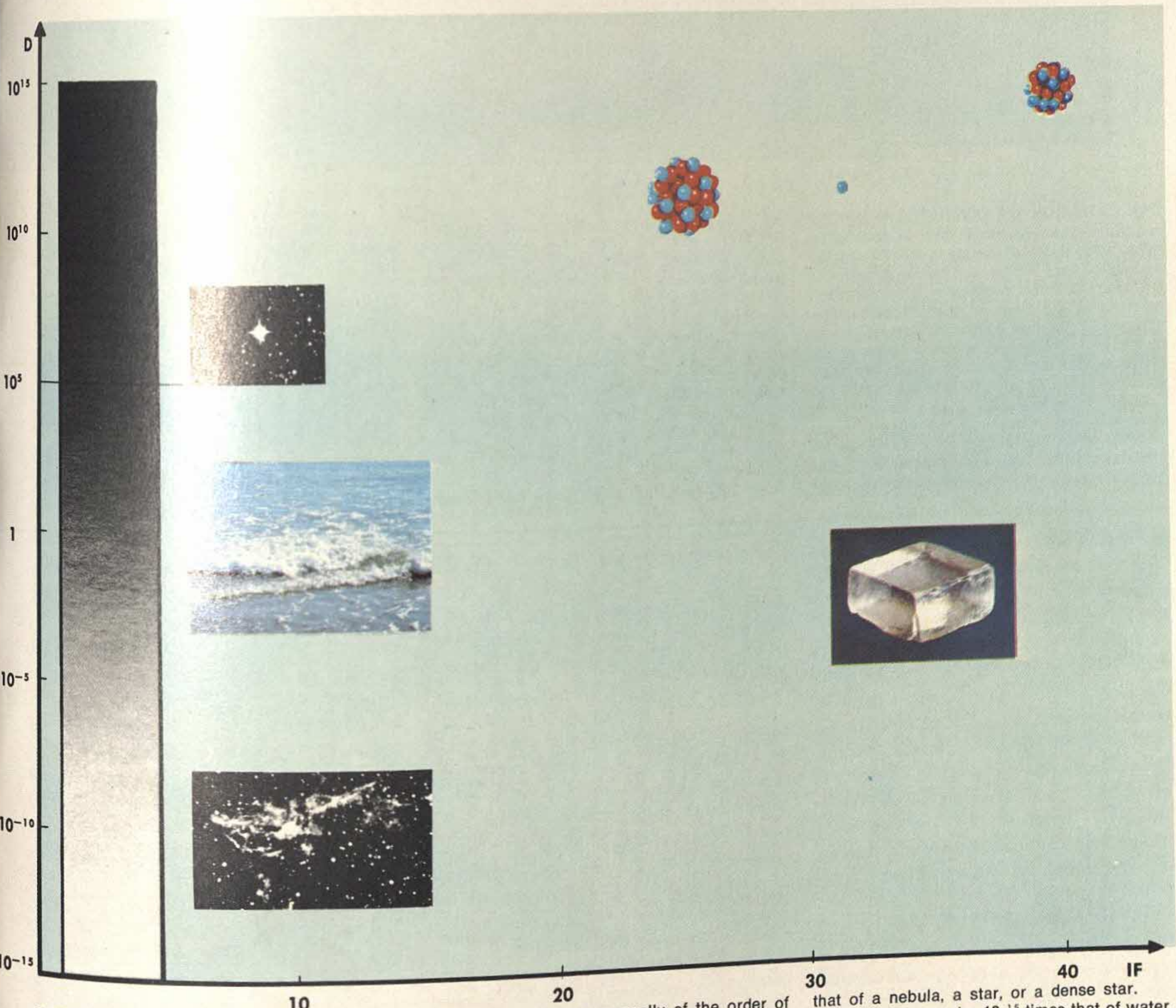
THE IMPOSSIBILITY OF SCREENING THE FORCE OF GRAVITY—While it has been established indirectly that the force of gravity is propagated at the same speed as light, direct proof could be obtained only if it were possible to screen the force of gravity. So far, such efforts have been unsuccessful, including the experiment illustrated here, in which two masses are placed on two plates at different heights on a balance. At the beginning of the experiment, the two masses are in equilibrium. A large mass of heavy material is then slid under the upper plate (the experiment has been tried with lead, mercury, and granite), but the presence of this screen of matter does not modify the equilibrium of the balance in any way, except that its mass attracts the weight immediately above it.

THE GRADATION OF FORCES

a revaluation of the force of gravity

Physics research has been extended to the astronomical field during the last decade because the laboratory of the universe offers far richer possibilities than any terrestrial laboratory. The consideration of stars of great density and stars of great mass has caused new theories to be put forward that call for many further astronomical observations before they can be confirmed.

The force of gravity has remained a great mystery for physics since the English scientist Sir Isaac Newton deduced its phenomenological laws. When addressing the problem of the force of



INTENSITY OF THE FORCES AND DENSITY— In this illustration, a scale has been constructed that shows the densities **D** of various structures, related to the intensity of the forces **IF** that keep them together.

It can be seen that matter in the universe becomes condensed and grouped into objects according to the types of forces that keep it united. It is these very forces that determine the density of the matter in each particular case.

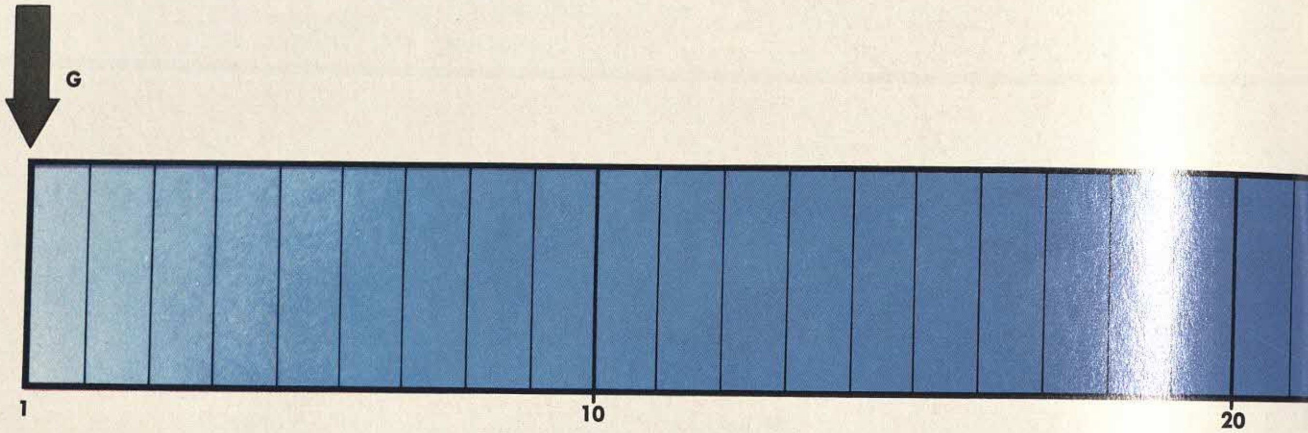
Very intense nuclear forces are capable of compressing matter to densities that are close to 10^{15} g/cm^2. Because electrostatic forces are less intense, they permit densities that are enormously smaller, generally of the order of the density of water; at the most, they can attain a value of about ten or twenty times the density of water.

The picture changes completely as regards the force of gravity. Until a short time ago, physicists had only studied the structures that can be obtained by means of masses of matter held together by this force in the limited range below a hundred thousand times the density of water, this being the maximum value of the density observed in the more common of the dense stars. However, the force of gravity tends to condense matter according to the structure into which the matter is arranged— that of a nebula, a star, or a dense star.

The density can be 10^{-15} times that of water, but can reach much greater values when the density itself comes into play to reinforce the gravitational field.

A rarefied nebula has little tendency to condense, but a star that is already dense (provided the conditions exist by which the atomic structures become weakened) will tend to condense even further.

Up to the present, the gravitational field has not been investigated above the limit marked by the horizontal portion of the illustration and little is known about extremely high densities.



THE CRITERION OF COMPARISON—All the equations that represent the variation of a force with distance contain a fraction in the numerator from which is read the product of the two quantities involved in the interaction (masses in the case of gravitational forces, charges in the case of electric forces, and so forth); the denominator shows a power of the distance that separates the interacting quantities.

The fraction is multiplied by a constant that translates the intensity of the force into the case in which the masses (or the charges, or other magnitudes) are unit masses and, moreover, are separated by a unit distance. For a comparison, a gram of hydrogen under normal conditions (that is, when it is biatomic), occupies a volume slightly greater than 10 liters (about 10 qts). About 0.00001 of this quantity is separated. Further, imagine that the electrons and protons of these atoms are separated. One coulomb of positive charge and an equal amount of negative charge will have been obtained. If these two charges are arranged (this is not possible in actual practice) at a distance of 1 m (about 3 ft) from each other, they would attract each other with a force of 10^9 tons. On the other hand, if two masses of 0.00001 g were placed at a distance of 1 m from each other, they would attract each other with a force that is enormously smaller. This force would, in fact, amount to no more than $10^{-[illegible]}$ tons, that is, 10^{-21} mg. The force of gravitation has never been taken into consideration as a possible interagent of the more intense electrostatic and nuclear forces.

The criterion for constructing a scale of these forces is provided by the dimension of the constant of the force in each particular case. This scale makes a comparison between the various forces possible.

The figures at the bottom of the scale indicate the values of the various forces. The force of gravitation is at the left-hand end of

gravity, Newton spoke his famous phrase *"Ipoteses non fingo,"* with which he indicated that he was interested only in how the force of gravity acted—not its origin or intimate nature. For over 150 years after Newton, the field of physics (specifically, mathematical physics) dealt increasingly with the laws of electromagnetic forces; that is, forces exerted between magnets or electric charges, or induced by the movements of charges or magnets. However, no theories were found that enabled physicists to develop a more profound interpretation of the force of gravity.

At the beginning of the present century a theory appeared which, at least in appearance, made a more profound analysis of the characteristics of gravity: the general theory of relativity. This theory was developed by the German-born physicist Albert Einstein. It demonstrated that the gravitational field deforms space in the immediate neighborhood of gravitating masses and that a straight line is no longer straight in the Euclidean sense. On the basis of this theory, subsequently verified by astronomical observations, it was also seen that light can change its wavelength in moving away from a strong gravitational field.

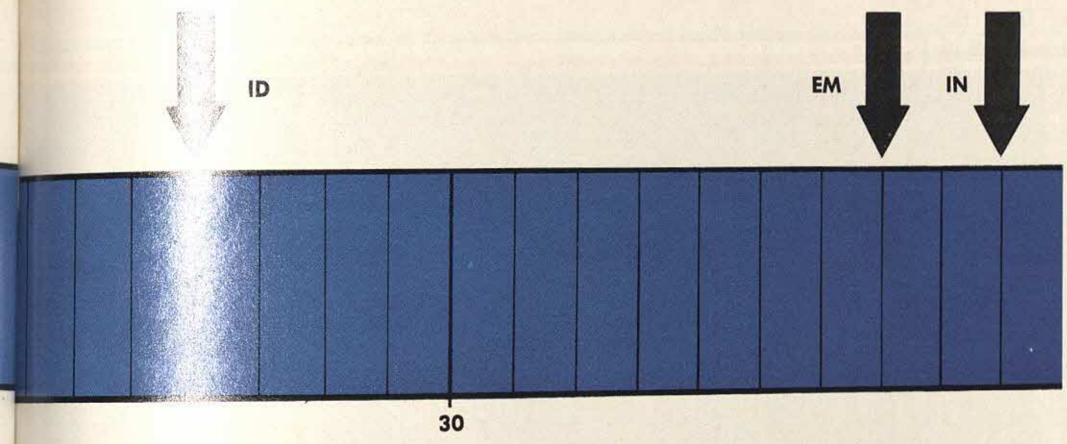
The force of gravity—which once had seemed to be as stable and untouchable as a principle of elementary geometry—was now seen to interact with other known fields; in this particular case, with the electromagnetic field. It was hoped that these laws of interaction might give a better insight into the nature of the force itself, but two factors intervened to delay progress in this direction.

The first was Einstein's unsuccessful attempt to construct a so-called unified theory of the electromagnetic and gravitational fields. He worked a great deal on this during the last years of his life, but succeeded only in giving a phenomenological and formal theory of the two fields. A real understanding of their nature was still lacking.

The second cause of delay was that physics soon became oriented toward the study of nuclear forces. Research began to move away from the classical problems of physics, which included the problem of the gravitational field. The discovery of forces whose intensity varied more rapidly than the inverse of the square of the distance brought physicists face to face with a new reality and with new mysteries. It no longer seemed feasible to reopen the problem of gravity when it was quite obvious that many other things remained to be understood.

It was also discovered that nucleons had an internal structure and that one had to imagine them as being made up of simpler units; moreover, these units must be held together by forces whose nature was as yet unknown, but which were likely to be different from known gravitational, electromagnetic, and nuclear forces.

This new discovery made it seem even more unlikely that an early solution might be found for the ancient problem of the force of gravity.



the scale: it is the weakest of all the forces and so is assigned a value of unity so that all the others may be described as multiples of it. Toward the right, but only after a long interval, is the force of the so-called weak interactions ID—that is, the forces that come into play in beta decay and affect the bond between electrons and nucleons inside the nucleons themselves. This force is as much as 10^{26} times greater than the constant of the force of gravity. The electromagnetic force **EM** is even greater, about 10^{37} times as great as the force of gravity. The force of the nuclear interactions **IN** is one hundred times greater than the electromagnetic force and is at the right-hand end of the scale. This scale has been left open because new forces might yet be discovered; for example, the force that binds the subnucleonic components and that is likely to be enormously greater than any of the forces so far considered. The force of gravity can be seen to be extremely small when compared with the other forces, to the extent that it was considered (with justification) to be negligible with respect to these others. This point of view is correct only when one considers quantities of matter that are either small or have a low density. The situation changes completely when these magnitudes assume greater values.

COMPARISON BETWEEN THE FORCES

During the period following the 1930s, a number of physicists (almost stealthily) continued to take an interest in the problem of gravity, particularly with a view to exploring the possibility that this force might be important in more areas than those encompassed by Newtonian thought. Some of these scientists began examinations of fundamental problems in physics and astrophysics that were to reach maturity only after the end of World War II. Problems of such dimensions are very closely connected with the possibility of carrying out appropriate observations. If, therefore, a physicist tackles them without having an opportunity to make any experiments, his work is only an interesting effort that is destined to remain sterile. Only when a number of research physicists dedicate themselves to studying both the theoretical and instrumental part of a problem are useful new insights gained.

For example, the penetrating preliminary analysis made by the American physicist J. Robert Oppenheimer in the investigation of neutron stars would have remained a piece of paper—and consequently fruitless—if it had not been for the subsequent discovery of pulsars.

It may well be that in the case of gravitational theories, only the study of quasars and pulsars will permit the discovery of some phenomenon that could lead to fruitful results in the practical field. These bizarre objects first detected in 1961 represent one of the fundamental astronomical discoveries of the twentieth century.

The initial considerations of the force of gravity were confined to only those cases involving enormous quantities of matter. Although a man's body also has a certain mass and exerts a gravitational pull on the body of another man, this attraction is so weak that it does not play any part in the movements made by either individual. The entire mass of the Earth must come into play before a man's mass is attracted with a force that is quite appreciable—that is, the weight of the individual.

A few decades ago, it was reasoned that if two neutrons are brought close together they are subject to gravitational attraction, but this attraction is negligible when compared with the nuclear forces that the two particles exert on each other. Gravity was therefore neglected. However, it was noted that under special conditions matter can produce gravitational fields so intense that the energy included and the other phenomena of motion become comparable to those caused by the forces of a stronger type.

Physicists, therefore, tried to construct a new theory of gravitational forces. They began to study their effect on matter under particular conditions and came to consider certain phenomena that require and suggest a new road for the study of the force of gravity. Among others, these phenomena include the so-called gravitational collapse of matter, or gravitational catastrophe—a phenomenon that implies nothing less than the disappearance of matter from the universe.

Such an event involves forces of presently inconceivable values. Its consideration exemplifies how the once-neglected force of gravity ultimately may be of greater importance than forces of an electromagnetic or nuclear nature. It is this direction of modern physics that makes the characteristics of the force of gravity of significance to the scientist.

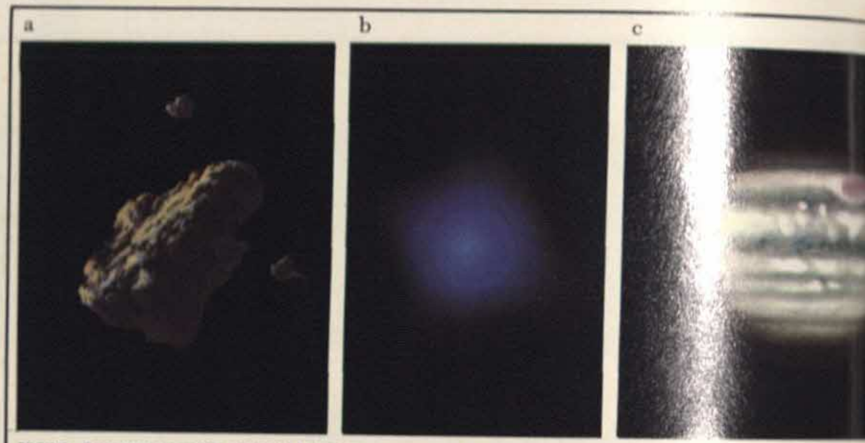
GRAVITATIONAL WAVES | the stability of all bodies in the universe

All bodies in the universe, from the planets down to the smallest particles of matter, have a mutual attraction for one another. Newton first demonstrated this in his law of universal gravitation. Subsequent experiments have shown that no matter what the circumstances, gravitation is always acting between two given bodies and cannot be destroyed under any known circumstances.

The gravitational constant is the gravitational force that a body of unit mass exerts on another body with the same mass at a given unit of distance from it. Nuclear forces are contained within a very small volume (the nucleus of an atom) and decrease rapidly a few diameters away from the nucleus. Electromagnetic forces have a much greater radius of action, but since not all matter is electromagnetic, this interaction is limited.

Gravitational forces are effective at any distance, but depend on the size (or mass) of the objects involved. In theory, the mass of a body that is constantly increasing in density will ultimately reach a state of gravitational collapse. In this

1



MASS, DIAMETER AND GRAVITATIONAL STABILITY—Several bodies with very different masses are shown in Illustration 1. The first (Illustration 1a) is quite small in the astronomical sense (an asteroid or meteorite), with a mass between a few and several hundred kilograms. In space, it exerts little gravitational force and can easily be fragmented. It would also be incapable of reuniting these fragments into a whole except at very low speeds. The pressure at the center of this body is very small, so much so that the constituent atoms are too solid to be condensed and rigid enough to withstand static gravitational pressure.

The second body (Illustration 1b) is a planet smaller than Jupiter (Illustration 1c) and subject to a great pressure at its center (millions of atmospheres) due to the enormous mass that weighs on the center. This planet, because of its large mass and considerable radius, is under great pressure from its gravitational attraction. However, because the body is cold and its atoms elastic, it can withstand the pressure and remain stable. Indeed, it is impossible to imagine a loss of stability in which these atoms would not be preserved indefinitely.

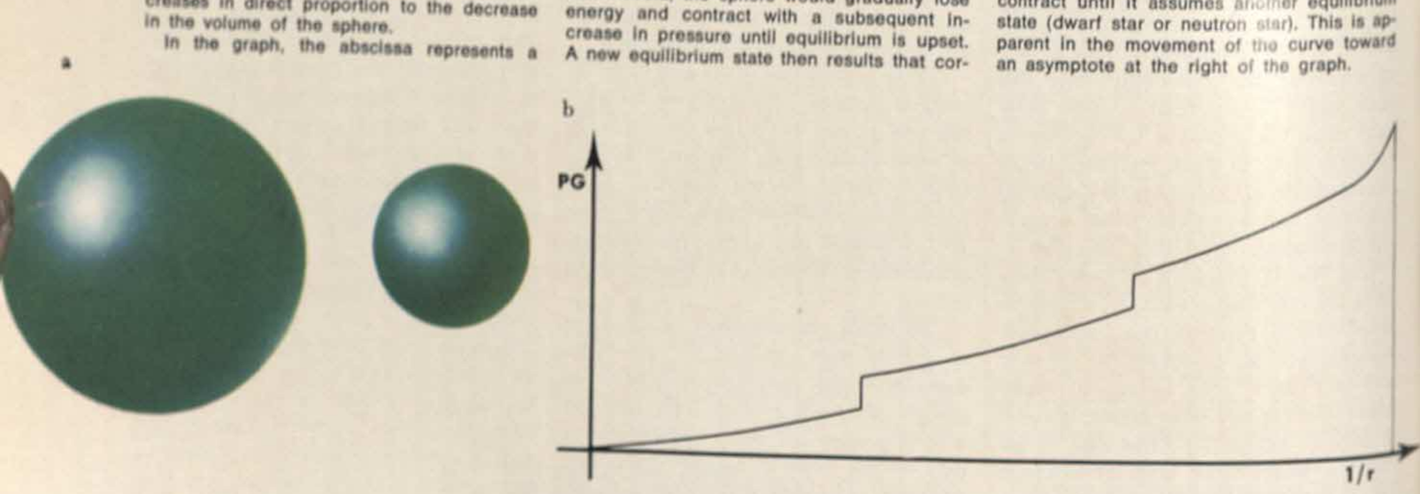
Finally, if a sphere the size of the sun were made to contract so that the nuclei of its

2

EVOLUTION OF GRAVITATIONAL STABILITY—A gravitating sphere with a definite mass and radius will exhibit a pressure at its center that can be determined. In fact, as a result of gravitational attraction, this pressure increases in direct proportion to the decrease in the volume of the sphere.

In the graph, the abscissa represents a magnitude that is the inverse of the radius of the sphere, while the ordinate represents a magnitude equal to the gravitational pressure at the center of the sphere. In the case of a nebula, the sphere would gradually lose energy and contract with a subsequent increase in pressure until equilibrium is upset. A new equilibrium state then results that corresponds to the first discontinuity in the graph.

For example, a nebula may pass from its present equilibrium state to that of a star, and the star may continue to lose energy and contract until it assumes another equilibrium state (dwarf star or neutron star). This is apparent in the movement of the curve toward an asymptote at the right of the graph.



atoms touched each other, the result would be a neutron star of extreme density (1,000 tons/in.^{3}). Such a star is, of course, subject to tremendously high pressure at its center. This pressure is great enough to destroy known atoms and the stability of the star depends instead on the rigidity of its nuclear structure (or nucleonic spheres). In this case, stability is also considered infinite since the rigidity of the nucleonic spheres is definite.

A dwarf star (illustration 1d) has a mass equal to that of the sun and a density 100,000 times greater than water. It is kept from collapsing because of the pressure at its center due to high-temperature stellar gases. The high temperature produces collisions between atoms of these gases, and the resulting pressure is enough to withstand gravitational attractions. This ongoing reaction is, however, quite unstable since the fuel can be exhausted. In this case, the dwarf star could collapse with another star being formed (probably a neutron star).

The stability of a normal star (illustration 1e) in the face of gravitational attractions is also due to pressures of thermal origin. Certain characteristics of the physical structure of such a star ensure that its density is less than that of a dwarf star although it is of the same mass.

Nebulas and galaxies (illustration 1f) are of considerably greater size than the various stars and have masses that vary from a few thousand to a few million times that of the sun. They avoid gravitational collapse because of their great size and because the various pressures (thermal, atomic, or nucleonic) are divided among stars and other bodies that constitute the nebula or galaxy.

Finally, a quasar (illustration 1g) has a mass far denser than any other celestial body. It is thus very close to a state of gravitational collapse, which may account for the unusually bright blue and ultraviolet light and the radio waves detectable from great distances.

state, the constituent atoms and molecules of the body would be incapable of further attraction. However, tremendous pressures would be required to bring about such a gravitational collapse.

The stability of all matter in the face of gravitational forces is of great importance in the study of physics. In his General Theory of Relativity, Einstein recognized the equivalence of gravitation and inertia, and he substituted a field concept for the Newtonian concept of gravitation as a force acting at a distance. Einstein's theory asserts that bodies produce curvatures in space that form a gravitational field, and that the path of a body in this field is determined by its curvature. Observations of starlight, made during eclipses of the sun, confirm this theory.

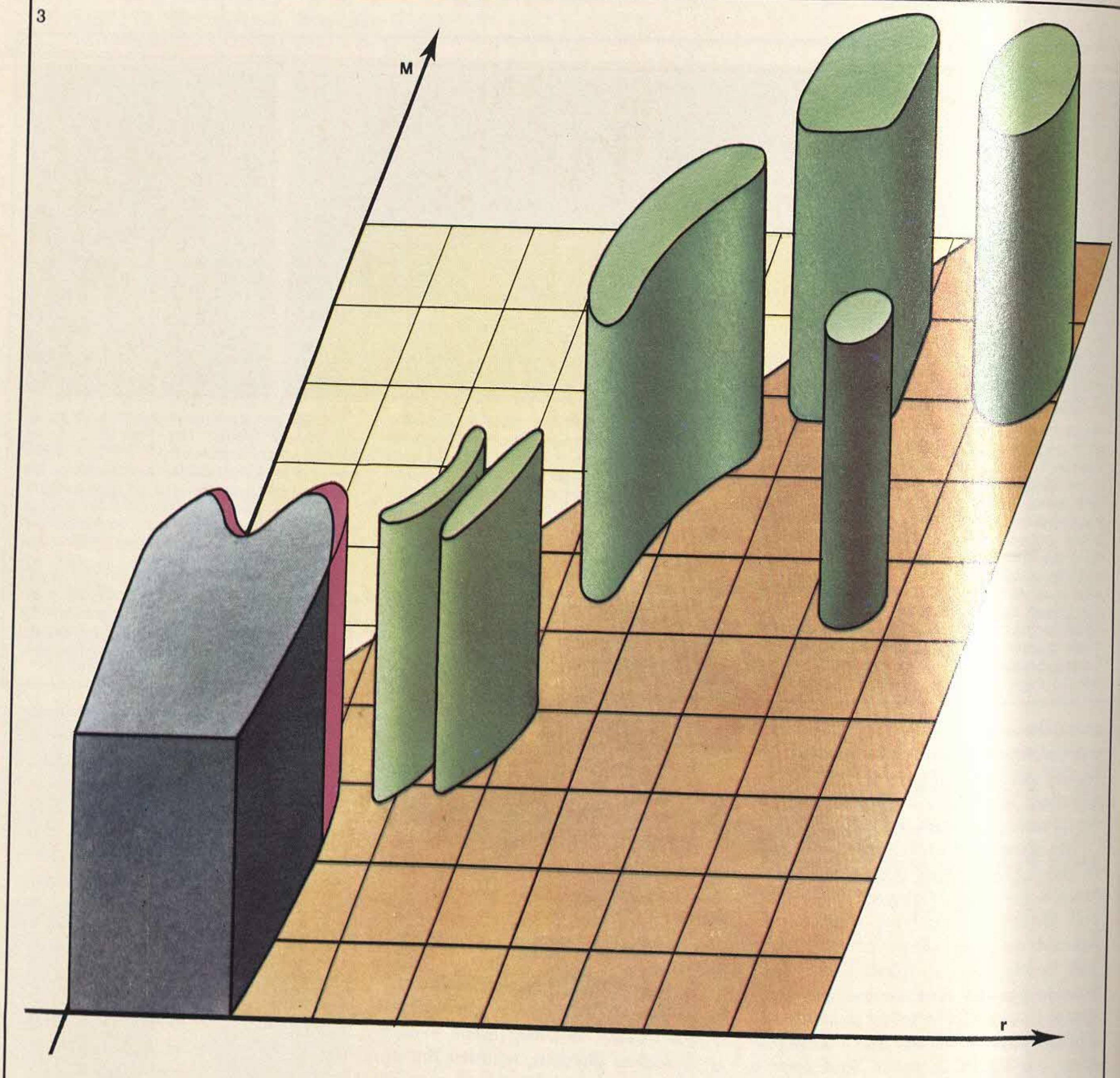
In his general theory, Einstein also postulated the existence of gravitational waves. He indicated that these waves should move through space (that is, through a vacuum) at the speed of light —3×10^8 m/sec (186,262 mi/sec). In this and many other ways, gravitational waves should be similar to electromagnetic waves. Any moving body should generate gravitational waves. However, in contrast with electromagnetic waves, gravitational waves are quite weak; therefore, only rapid changes in the shape of an extremely massive body would produce detectable gravitational radiation.

For a long time, no scientist was able to confirm the existence of gravitational waves. Finally, in 1969, Joseph Weber, an American physicist, reported that he had actually observed this kind of radiation. Weber used specially constructed equipment consisting of a large aluminum cylinder suspended horizontally in a vacuum chamber by a thin wire. The cylinder, five feet in length, had a diameter of 26 inches. The apparatus was carefully constructed to shield it from disturbances of electromagnetic and seismic natures. Whenever a gravitational wave or wave pulse struck the cylinder, it caused the cylinder to oscillate. Special transducers permitted Weber to detect this oscillation.

Weber actually used two detectors, located hundreds of miles apart. He found that a substantial number of wave pulses struck the two detectors simultaneously; the number of pulses far exceeded the probability of accidental coincidence.

Study of Weber's results indicates that gravitational waves are being emitted by a source or sources in the Milky Way—probably in or near the center of the galaxy. As yet, the exact sources have not been identified. Some scientists have suggested that the sources of these waves are spiraling neutron stars or pairs of neutron stars. Many scientists hold that the powerful bursts of gravitational waves observed by Weber could emanate only from the gravitational collapse of a star, which results in the formation of a so-called black hole, or from the collision of two black holes.

3



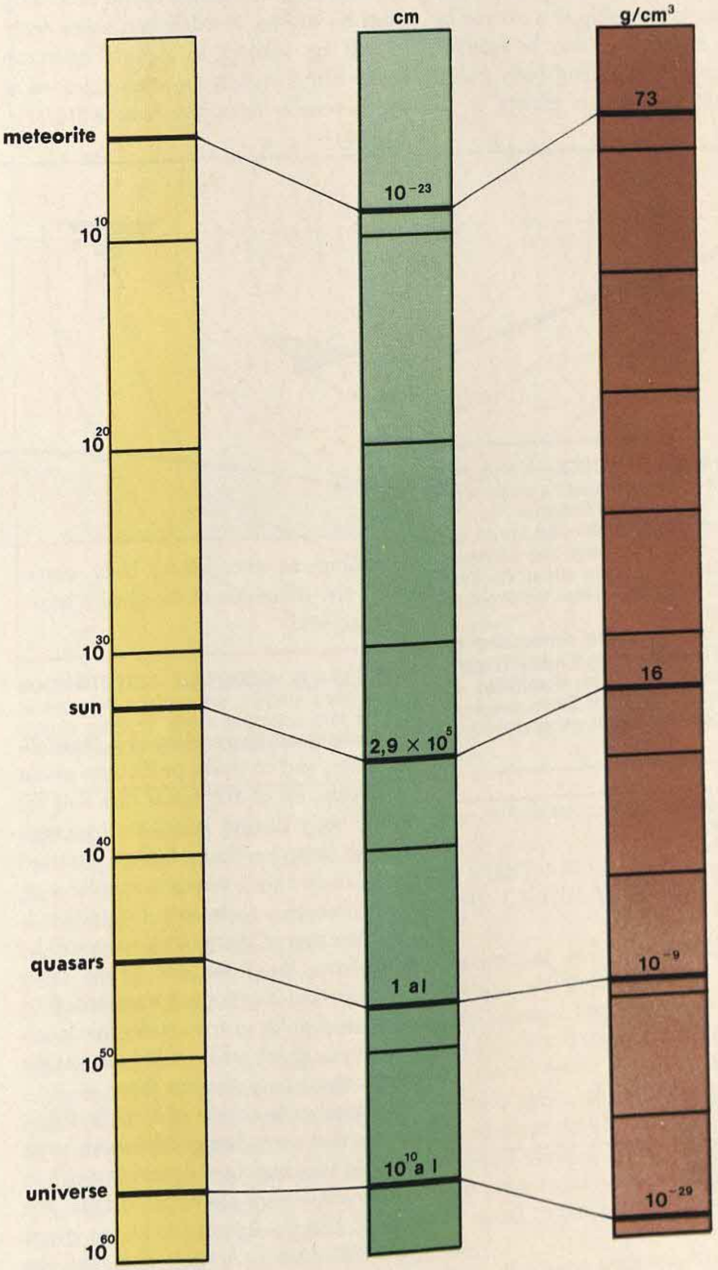
GRAVITATIONAL STABILITY OF HEAVENLY BODIES—The radius and mass of several celestial bodies can be used to illustrate their relative gravitational stability. The bases of the solids in this case enclose points that correspond to bodies that really exist in the universe. These areas have been raised (made solid) to show up more clearly; the oblique line at the bottom right is that area in which the bodies are considered stable.

This stability is the result of the forces mentioned (elasticity of atoms, rigidity of nuclei, thermal or mechanical agitation, and so forth). Those bodies with a very small radius and large density are those incapable of withstanding gravitational attractions; such bodies continue to contract until new equilibrium states are reached.

THE GRAVITATIONAL RADIUS—Any concentration of matter that is dense enough will have its atoms or nuclei in contact with each other so that further attractions may cause a gravitational collapse. In this case, the gravitational attraction among the components will be sufficient to overcome any inhibiting properties. The three scales represent the masses of various bodies, the maximum diameter to which they must contract to reach a state of gravitational collapse, and the density they must attain when the sphere radius equals the gravitational radius.

For example, a meteorite the size of a human body must contract until it has a radius of 10^{-23} cm and a density of 70^{73} g/cm³ before gravitational collapse results. No situation is known where such spontaneous contraction could occur. In the case of the sun, it is necessary to reach a diameter of 2.9 km and a density of 10^{16} g/cm³ (close to that of neutron stars) before gravitational collapse takes place. Conversely, the entire universe is separated from this state by a factor of no more than ten—much less than that of a meteorite, the sun, or a distant quasar! Thus, gravitational collapse of the universe and all matter could conceivably take place in ten million years.

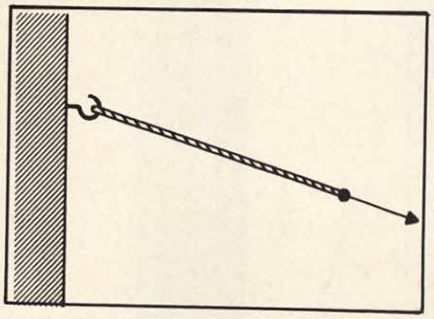
4



FORCE FIELDS | weight, gravity, and falling bodies

To understand how an object is acted upon by a force, the nature of the force must be known. The force may come from within, as in a ship or a car run by an engine, or the force may be external, as in the case of a falling body pulled earthward by the force of gravity.

1

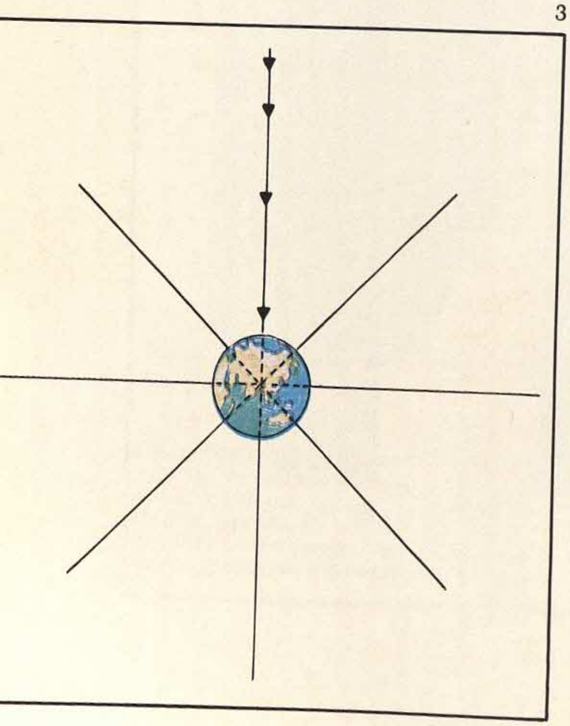


DIFFERING FIELDS OF FORCE—A rope pulling on a hook is an example of a single force concentrated at a point (Illustration 1).

The force of attraction directed toward the center of the Earth is a force that decreases with an increase in altitude above the Earth (Illustration 2). The arrows show the directions that such force takes.

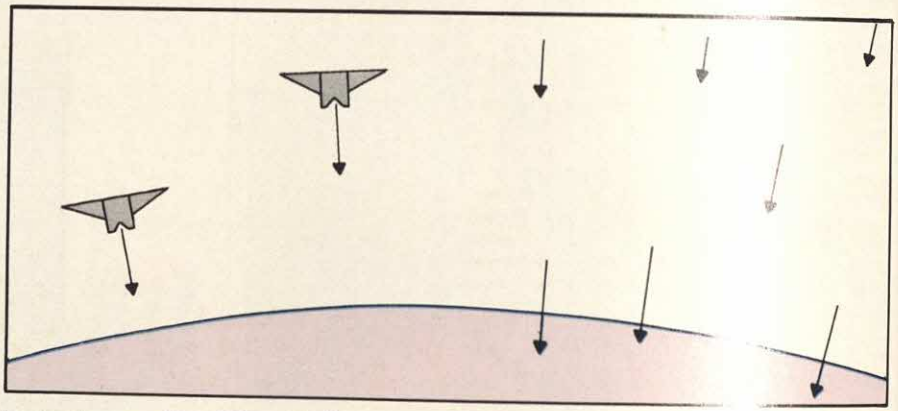
At every point in space surrounding the Earth, a mass is acted on by the Earth's gravitational force of attraction (Illustration 3). Those forces far from the Earth are small (short arrow); those closer are greater (long arrow).

3



To predict the behavior of a free-falling body, all the forces to which it will be subjected during the course of its fall must be known. If it falls only a few feet, it will be subject to a fairly constant force. But if it falls for many miles, as a space vehicle does, the force will vary, increasing as the falling body comes under the influence of the Earth's gravitational field.

2



FIELDS OF FORCE

To understand the motion of a free-falling body, and to make predictions about its motion, all of the forces that will act on the body during its fall must be considered. Suppose that a hook is attached to a wall and that someone is pulling on the hook with a rope, as in Illustration 1. The direction of this pulling force will be away from the hook and in the same direction that the rope is stretched. The problem would automatically be localized at one point and would have to take into account only the one force.

As another example of a single force, assume that some heavy body, such as an anvil, is attached to a dynamometer (an instrument which shows what force the anvil is being subjected to at any given point in space). Assume also that the anvil is near sea level and that the dynamometer shows that the anvil weighs 20 kg (about 44 lb). It is then obvious that a 20-kg force is pulling the anvil toward the center of the Earth. This force can be represented by an arrow whose length is proportional to the force, and that points from the middle of the anvil toward the center of the Earth.

Now suppose the same anvil is moved to a position that is 100 km (about 62 mi) above sea level. The dynamometer will show that the anvil now weighs about 0.5 kg (about 1 lb) less than it did at sea level. The weight-force is still directed toward the Earth's center. An arrow indicating the direction of force at this new location will not be parallel to the first arrow; rather, both arrows will point toward the Earth's center, as in Illustration 2.

These two tests suggest that at any point in space around the Earth a material body is subject to a force, called weight, which is caused by the Earth's gravitational pull on the object, and that the force decreases as altitude increases. More precise tests would lead to the conclusion that there is a field of force around the Earth that exerts a pull on a physical body that is directly proportional to the mass of the body (the quantity of matter in it), and inversely proportional to the square of the distance of the body from the center of the Earth. A chart could be made showing the weight-force that will act on a solid object at any point around the Earth.

LINES OF FORCE

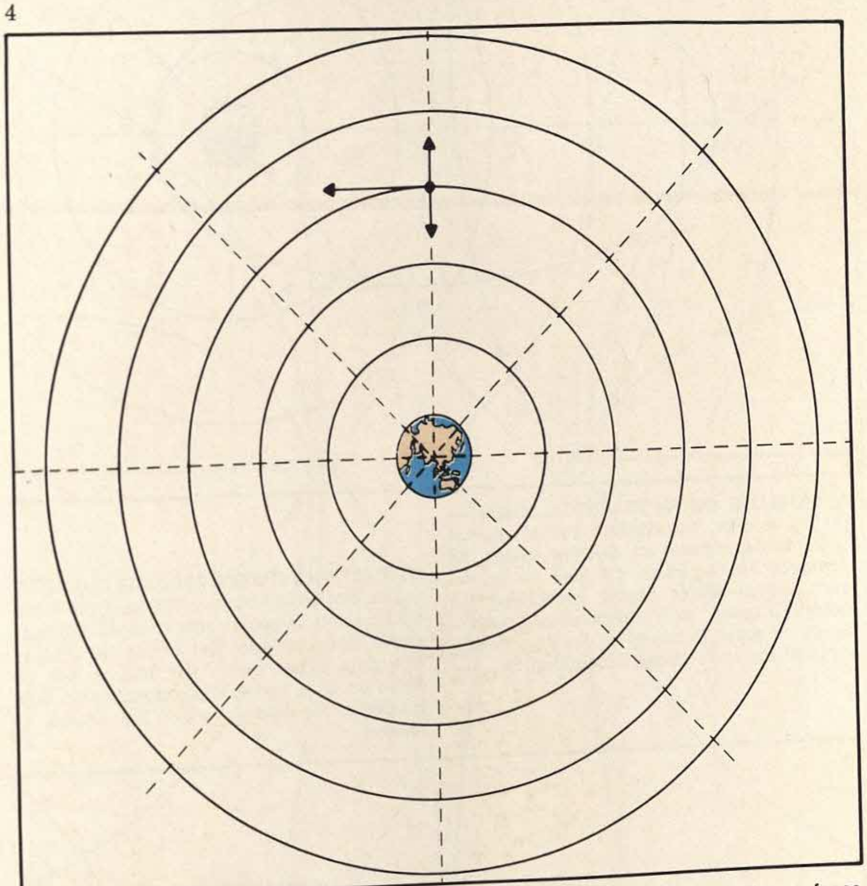
Illustration 3 represents the Earth as seen from space. Its geographic North Pole is at the top, its South Pole at the bottom. Imagine being situated at a point on the imaginary projection of the vertical axis above the North Pole, at a distance many times greater than the Earth's diameter. Then imagine drawing, down along the axis, an arrow representing the weight of an object at that point. Next, imagine moving down the projected axis to the tip of the arrow. The weight of the object will be greater than it was originally. A new arrow, representing the new weight of the object, would be longer than the first arrow.

Then, moving on to the point of the new arrow—and finding the weight of the object increased still more—a longer arrow can be drawn. One could continue in this fashion until reaching the surface of the Earth and the series of arrows would form a straight line from the original position to the Earth's surface. This straight line represents a *line of force.* For the Earth's gravitational field of force, all the lines of force are straight lines converging at the center of the Earth. (In many other situations, however, the lines of force are not straight; such lines could be studied in the same way, but for now it is only necessary to be concerned with those essential to understanding the forces around the Earth.)

EQUIPOTENTIAL SURFACES

As everyone knows, a car traveling on a flat road must continually use gasoline. If it did not have to overcome road friction and wind resistance, it could simply coast on and on. When the car travels uphill, it uses even more gasoline because the engine is working against gravity. Going downhill, the car's engine will use less gasoline than on a level road because gravitational field forces are working in the car's favor.

4



EQUIPOTENTIAL SURFACES—In any field of force, each equipotential surface possesses a unique property: every point on that surface has the same potential. Different surfaces have different potentials, or forces. The surfaces are shown concentric with the Earth's sphere. The force of gravity acts equally on any point of a given equipotential surface.

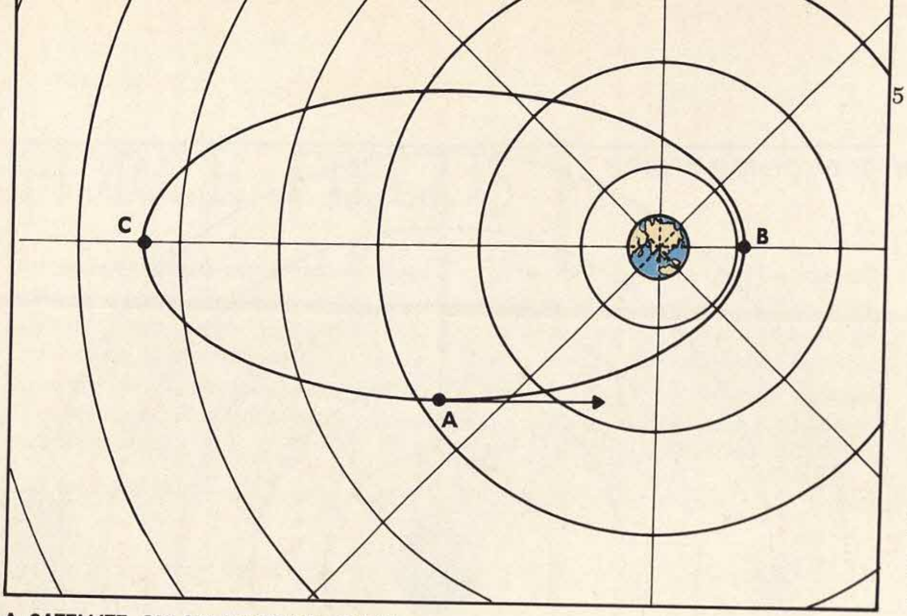
From the car analogy, it can be seen that whenever a body that has weight moves along a spherical surface concentric to the surface of the Earth it neither gains nor loses energy because of gravitational field forces. Illustration 4 shows some spherical surfaces around the Earth along which a body can move without losing or gaining energy. These spherical surfaces are called equipotential surfaces of the field.

From all points on any one of these spheres, a body dropped toward the Earth would acquire an identical quantity of energy during its fall. Therefore, at all points on a spherical surface concentric to the Earth's surface, a body possesses the same "potential energy" before it falls toward the Earth.

A FIRST APPLICATION: THE CIRCULAR ORBIT

Suppose an object is propelled into the Earth's gravitational field (or into the gravitational field of any celestial body) with a specific velocity perpendicular to a line of force. If the velocity is sufficiently great, the object will achieve a permanent circular path concentric to the Earth's surface. This takes place when the centrifugal force acting on the object is equal to the gravitational force acting on it.

In the case outlined, the object will continue eternally in its circular orbit as long as no other forces intervene. Since



5

A SATELLITE ON AN ELLIPTICAL ORBIT—During its orbit, the satellite crosses several equipotential surfaces of differing values. As it approaches the Earth, the field forces impart positive kinetic energy, increasing the satellite's speed. As it travels away from the Earth, its speed is slowed by the negative effects of the same forces (Illustration 5).

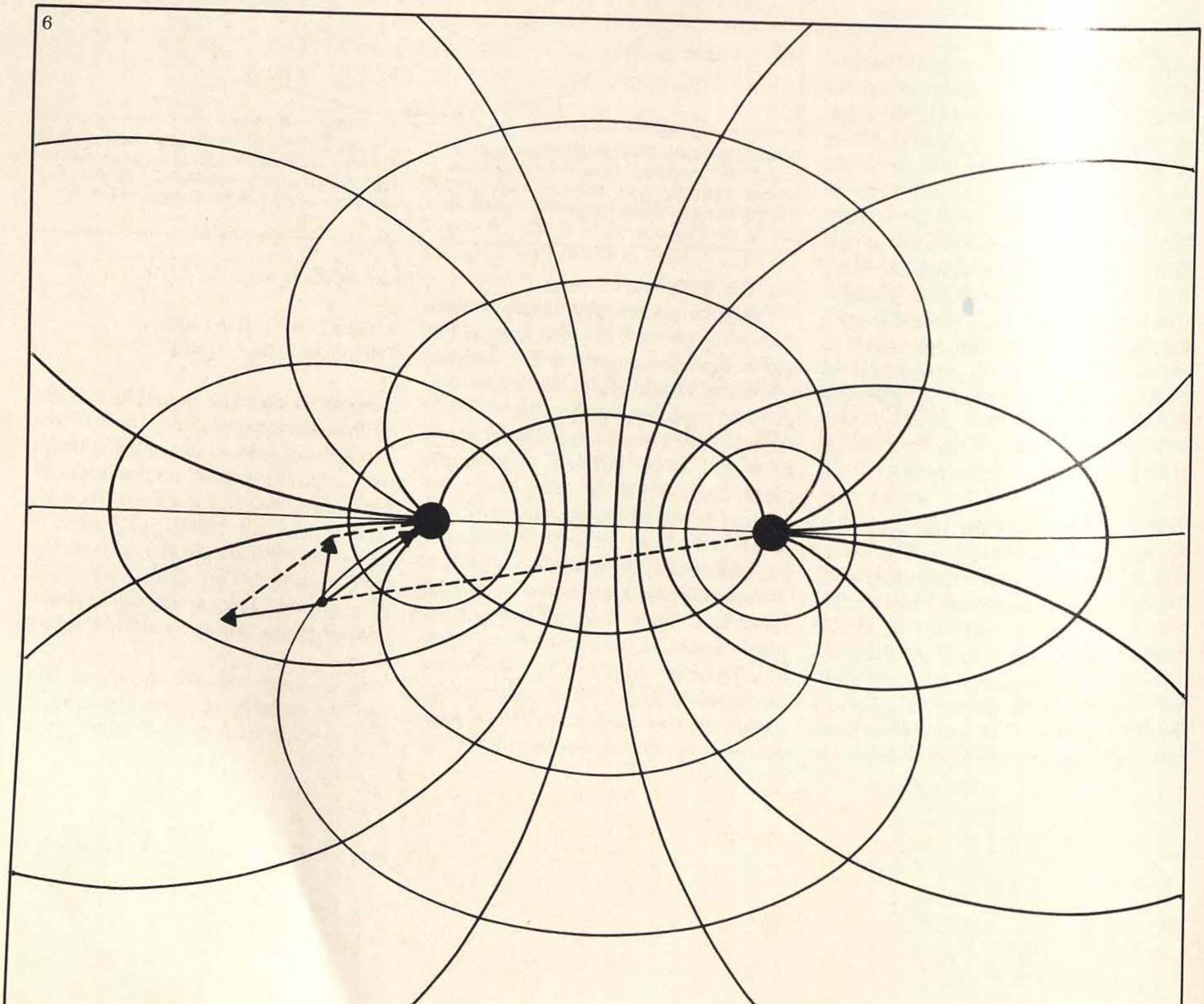
force-weight is always exactly perpendicular to the object's circular orbit, it is a constant force that cannot accelerate nor slow down the body tangentially.

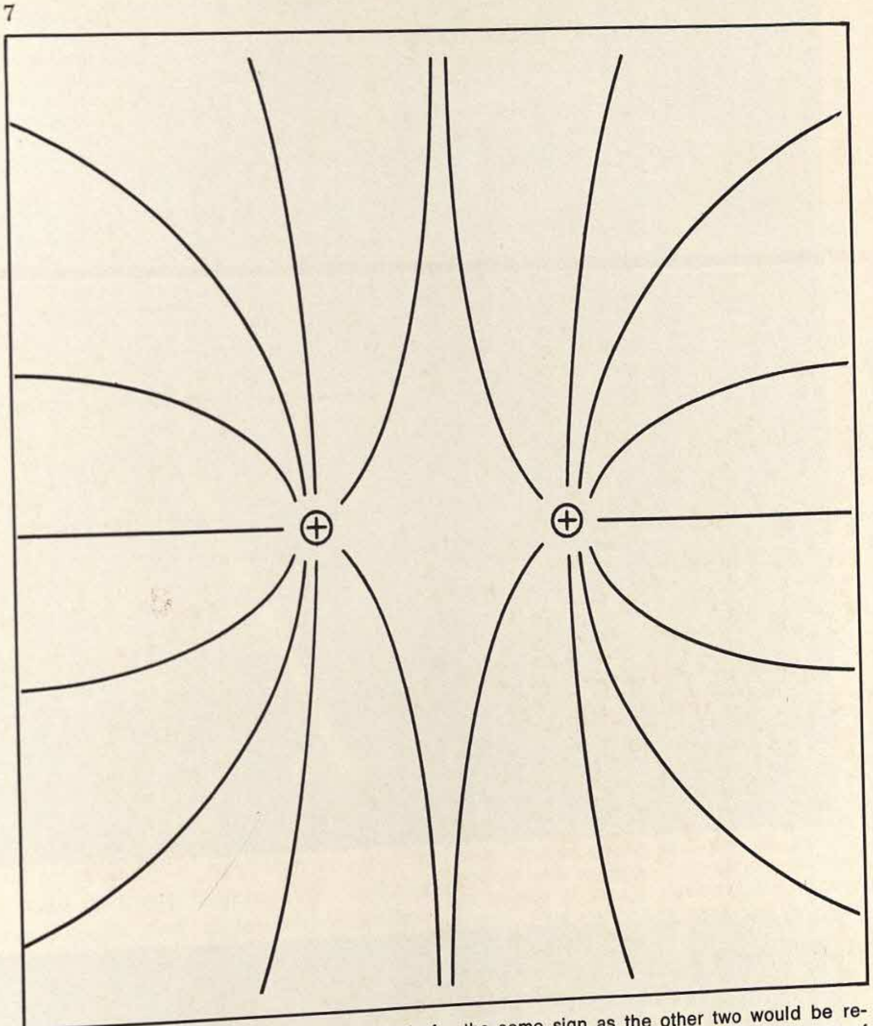
A SECOND APPLICATION: THE ELLIPTICAL ORBIT

Illustration 5 shows the elliptical path of a body moving in orbit around the Earth. Through knowledge of the nature of the Earth's gravitational field, the behavior of this body, as it moves along its orbit, can be predicted. As it approaches the Earth, moving from C to B and passing through **A**, the body crosses several equipotential surfaces. The field consequently exerts a force in the direction of the body's motion, thus increasing its speed. The speed, which was smallest at point

ATTRACTION BETWEEN OPPOSITE CHARGES—This diagram (Illustration 6) shows the force field created by two electric charges with opposite signs but identical values. If another charge is placed within the field, it will be acted on by a force whose direction is tangential to the line on which the charge is located.

6





REPULSION—This is an example of a field of force due to two charges of the same sign and of the same intensity. A third charge of the same sign as the other two would be repelled along similar lines; a third charge of opposite sign would tend to be attracted.

C, becomes greater at point **B.**

As it moves from B back to **C,** the body again crosses the equipotential surfaces, but in reverse order. In this portion of its orbit, the body will slow down because the field's forces are acting in a direction opposite to its motion. It is possible to describe with mathematical formulas the shape of the lines of force and the equipotential surfaces of the Earth's field. Using these formulas, it is possible to predict the exact shape of the body's orbit within the gravitational field of the Earth. This kind of calculation is, in fact, used by scientists in determining the orbits of planets and satellites.

A MORE DIFFICULT CASE: THE DIPOLE

In addition to these force fields caused by gravitation, there are many other kinds. One example is the force created by an electric field. An electrical charge has around it a field of force that will attract unlike charges and repel like charges. The force field around an isolated electric charge would look similar to Illustration 4.

Suppose two electric charges of equal strength—one positive and one negative—are placed next to each other. If a third charge is placed in the space surrounding these two charges, it will be subjected to a force that is similar to that shown in Illustration 6. An atom that has lost an electron—and the electron itself, if not too distant from the atom—will form a field with lines of force like those shown in Illustration 6.

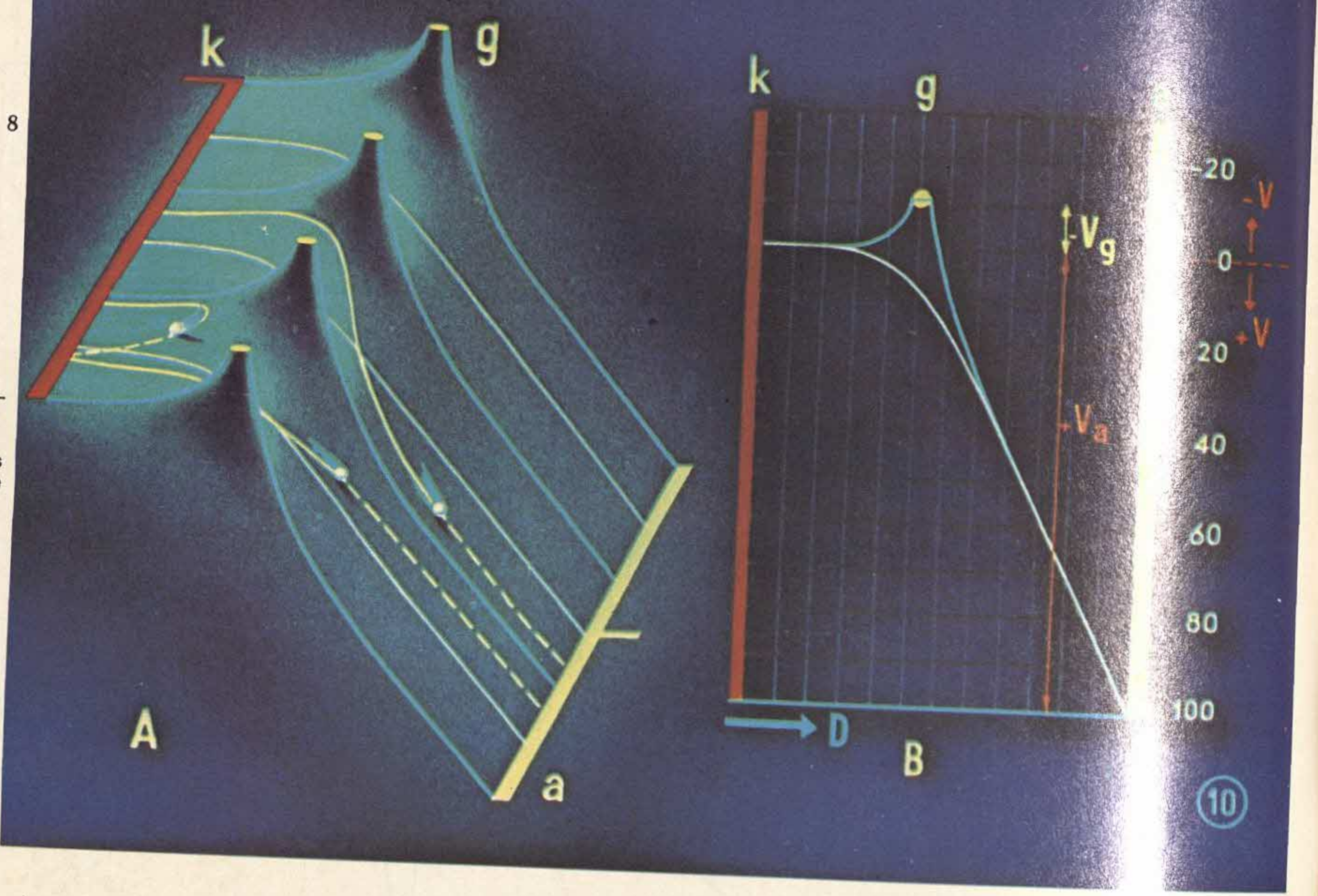
Illustration 7 provides another graphic example of a field formed by two centers of force. In this case, the two charged particles possess the same positive charge. If a third charge with the same sign is placed on any of the lines of force, it will be repelled by both charges, and will move outward along the line of force, while a charge with the opposite sign will be attracted toward the charges and will move inward along the line of force.

A PRACTICAL CASE

It has been established that knowledge of a field's forces will allow the motion of bodies within the field to be predicted. To diagram the motion of a space probe traveling toward the moon, for instance, it must be known to what extent the Earth's gravitational field is changed by the moon's. This calculation is complicated by the fact that the moon is in motion with respect to the Earth. At a given point in space around the Earth, therefore, the forces exerted by the moon and the Earth change continuously.

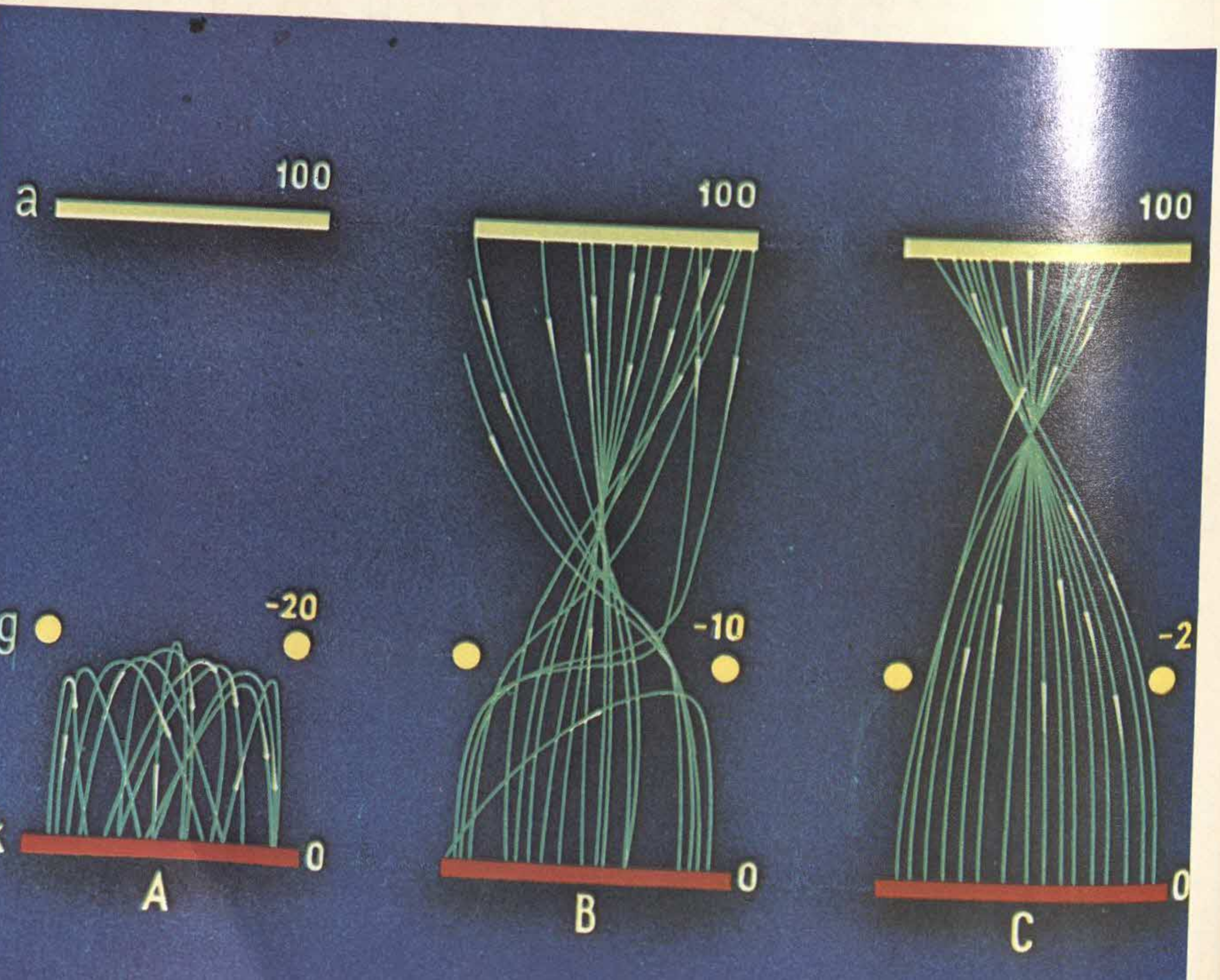
Another practical example, shown in Illustrations 8 and 9, is an electron traveling through a thermionic (vacuum) tube —in this case a triode, a type of tube often used as an amplifier in electronic circuits. The triode is composed of a cylindrical metal sleeve from which electrons are emitted. The electrons pass through a surrounding concentric wire-mesh cylinder that attracts them and are then collected on a metallic plate. By knowing the shape of the force fields to which an electron in this system will be subjected, the geometrical characteristics of its trajectory, as well as the peculiarities of its movement, can be predicted.

8



THE TRIODE— Fields of force are applied in many branches of science. The pattern of forces on an electron inside a thermionic, or vacuum, tube is shown. The letters **a, g,** and **k** represent, respectively, the anode, the grid, and the cathode. Illustration 9 represents the ways the potential of the grid varies the field of force with the relative line of force. In the event in which the potential is −20, no electron emitted from the cathode will pass through the mesh of the grid. When the potential is −2, the electrons emitted from the cathode reach the anode and always maintain a value that is positive with respect to the cathode.

9



NUCLEAR REACTIONS | from low to high energies

An important branch of physics is the study of the structure of the atomic nucleus through observation of the reactions to which it may be subjected: in other words, nuclear reactions. By nuclear reaction is meant the response of an atomic nucleus when it is struck by a particle with a certain kinetic energy. Usually, another particle or the same particle is emitted, or energy is emitted in the form of electromagnetic radiations such as gamma rays.

Chemists describe the atomic nucleus simply as that part of the atom in which are concentrated almost all of its mass and the positive charge that keeps the electrons in their various orbits. The nucleus is composed of neutrons and protons; its total charge is the sum of the single charges of its protons, and its mass is approximately the sum of the masses of its protons and neutrons.

Because nuclear reactions are extremely complex and varied, physicists try to cause a vast range of such reactions by bombarding all kinds of nuclei with many different kinds of particles with all the energy values to which the particles can be accelerated. Some of these particles are protons, neutrons, and in some cases electrons, deuterons, alpha particles, or electromagnetic radiation. The range of the energies of the particles used is very extensive, from thermal energies up to the high energies produced by large accelerators.

It is in fact possible to induce both atomic and nuclear reactions by impact between particles having energies of the order of the kinetic energy of an atom at ordinary temperature. An example of the former is the case of oxygen and iron atoms—mere contact of the two atoms is practically sufficient to bring about a reaction of combination. In the same way, the equally low approach energy of a neutron is sufficient to bring about a transformation of the nucleus it meets; that is, to produce a nuclear reaction. In these instances, the energies released are of the order of only a few electron volts.

On the other hand, at energies exceeding 1 Bev (one billion electron volts), a particle striking the nucleus will cause most of its changes in the nucleons that compose the nucleus. In this range of energies, the collision between the bombarding particle and the nucleons of the nucleus occurs as if the nucleons were in a small, loosely bound cluster, because the binding energies of the nucleons—the energies that keep them together to form the nucleus—are so minute in comparison with the energies of the collision itself as to be insignificant. This range of energies may be called the range of subnuclear reactions. The reactions occurring in this range provide information about such elementary particles as the nucleons themselves, rather than about the structure of the nuclei. Various kinds of mesons and so-called strange particles are produced at very high energies in these collisions; therefore, this range of energies is used to investigate mesonic phenomena.

In 1935 the Japanese physicist Hideki Yukawa predicted the existence of mesons. They were detected by U.S. physicists who were making high-altitude studies of cosmic radiation in 1937. Meson theory became an important part of nuclear and high-energy physics. Mesons take their name from the Greek *mesos*, meaning middle, because they have masses that are intermediate between those of electrons and protons. Since their discovery in cosmic radiation, mesons have been produced by high-energy particle accelerators. A considerable variety of these particles has been discovered; some are neutral, while others carry a unit of positive or negative charge. They differ not only in their charge but also in their masses, spins, lifetimes, and modes of decay. The most common mesons are pi-mesons and K mesons. At least the pi-mesons (one class) are closely connected with the forces that bind the nucleons into stable nuclei.

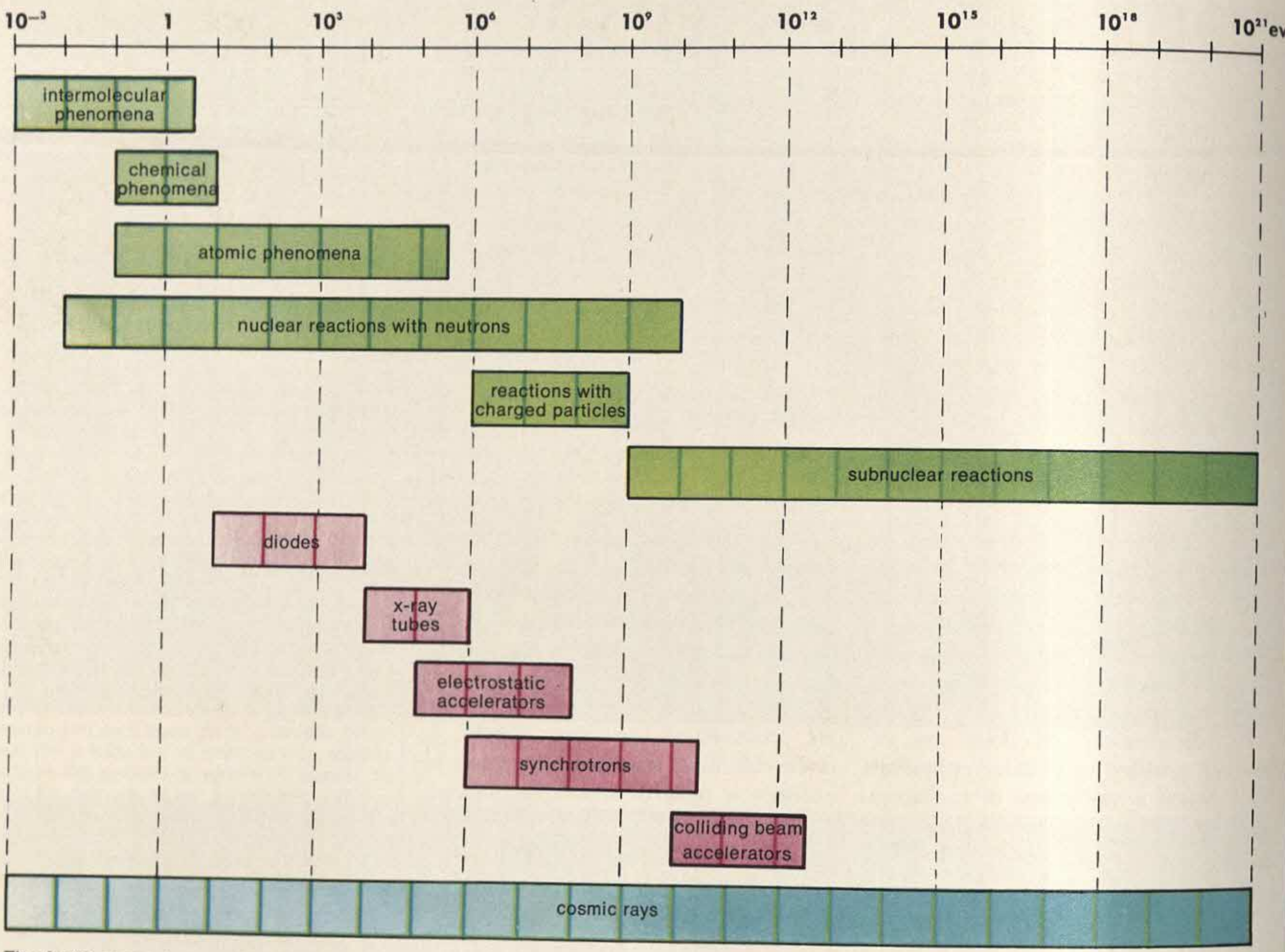
INTERMOLECULAR, MOLECULAR, ATOMIC, NUCLEAR, AND SUBNUCLEAR PHENOMENA —This table summarizes the energies at which the principal categories of physical phenomena take place.

The scale at the top of the table covers the entire range of energies involved in this kind of phenomena, in electron volts. The green bars below this scale indicate the range of energies within which each category of phenomenon occurs. The pink bars represent the energy ranges of the more familiar kinds of energy-producing devices, including accelerators. The blue bar at the bottom represents the energy range of cosmic rays.

The green bar at the top left corner represents the category of phenomena that may be designated as intermolecular; these phenomena involve the least amount of energy. If some atoms or molecules are placed near each other, they manifest forces of reciprocal attraction that keep them together to form solids or liquids. These same forces are also the forces important in the collisions between the molecules of a gas. Such forces are rather weak, and range from 0.01 ev to little more than 1 ev. If, for example, a particle with a kinetic energy of many electron volts is propelled against a solid, each time this particle collides with an atom of the solid it will lose an amount of energy exceeding the binding energy of the atom it has struck, and the atom will be displaced. Even though the collision is local, the entire crystalline structure of the solid is destroyed. The weakest interaction between an accelerated particle and matter thus ruptures the very structure of the matter.

Chemical phenomena represented in the second green bar involve the energies developed in the attraction that takes place between atoms—the so-called binding forces. These vary from a fraction of an electron volt to a maximum of a few electron volts. This range of energies is far below that involved in the more common nuclear reactions.

A somewhat higher and more extensive range of energies is that involved in atomic phenomena (third green bar), those involving the bond between the electrons and the nucleus of the atom. The emission and absorption of light, ultraviolet radiations, x-rays, and the ionization of atoms in a plasma as a result of collision—all belong within this range of energies. At one end of the range, very small energy exchanges, of approximately 0.1 ev, may be sufficient to cause the emission of electromagnetic radiations; at the other end of the range, energies of more than 100,000 ev (10^5 ev) are needed to extract the innermost electrons from atoms. A particle accelerated to an energy of many million electron volts, therefore, not only breaks the bonds between atoms, but also destroys the electronic structure of the atoms themselves.



The fourth green bar represents the range of nuclear reactions caused by neutron collisions. Such reactions take place over a very wide range of energies, from a fraction of an electron volt (a neutron with such energy is called a thermal neutron) up to about 10^9 ev (1 Bev).

The next, very short bar represents nuclear reactions with charged particles having sufficient energy to overcome the repulsion of the nucleus. Electrons, because of their very small mass, are not particularly suitable for the production of nuclear reactions; most charged particles (such as protons and deuterons) used for nuclear reactions are all positively charged. Before colliding with the nucleus, such charged particles must overcome the repulsion caused by the positive charge of the nucleus. The intensity of this repulsion depends on the charge of the propelled particle, the charge of the nucleus about to be bombarded, and certain additional properties of that nucleus. In very rough approximation, a particle with a charge z about to strike a nucleus with an atomic number Z and an atomic weight A must have an energy of about $zZa^{1/3}$ Mev. In general, an energy of about 1 Mev is necessary in order to produce a nuclear reaction with a charged particle.

The last of the green bars, with a very extensive energy range above 10^9 ev (1 Bev), represents interactions among subnuclear particles. At this energy level it is of little importance whether the particle struck is or is not part of an atomic nucleus. For the same reason it is of no importance, in studying collisions in this energy range, whether the target nucleus is surrounded by all or only some of the electrons indicated by its atomic number, or whether the atom does or does not form part of a solid. The binding energies of the electrons as well as those linking the atoms of a solid are of a magnitude that neither contributes to the collision phenomena occurring between nucleons nor limits these phenomena; and this applies also to the binding energies that link the target nucleon to the other nucleons in the nucleus. The upper limit of this range of energies is very high, as much as 10^{21} ev (1 trillion Bev), corresponding to the energy of a bullet fired from a small-caliber gun. Particles with such high amounts of energy cannot be produced in the laboratory, but they are found in cosmic radiation.

The pink bars indicate the energy ranges of some man-made devices used in the laboratory. The first bar represents diodes, everyday thermionic tubes, in which the energy level ranges from about 10 ev to a few thousand electron volts or, in very special cases, above 10,000 ev. The range of energies to which electrons must be accelerated to produce x-rays from an anticathode extends from 10,000 to 1,000,000 ev (1 Mev). The most powerful of the accelerators so far constructed provide energies of about 30 Bev. Energies greater than this, up to several hundred Bev, will be obtained by means of even larger accelerators now being designed.

The blue bar at the bottom represents the energy range of cosmic rays. The left portion represents the energies of cosmic particles found in the so-called secondary spectrum, the spectrum of rays that can be observed within the atmosphere. The right portion represents the primary spectrum, the rays that arrive from space. Because of the vast range of the energy spectrum of cosmic rays, it is easy to understand why physicists always had recourse to these rays when they sought to produce nuclear reactions before high-energy accelerators were available. However, accelerators are now generally preferred because they make it possible to carry out experiments more rapidly and under controlled conditions. Nevertheless, as science continues its exploration of reactions that take place at the highest energy levels, cosmic radiation must still be used.

THE PRODUCTION OF ARTIFICIAL NUCLEI | completing the periodic table of elements

When the Russian chemist Dmitri Mendeleev organized his system for tabling the known elements by their ascending atomic weights, it became apparent that the table contained a number of gaps. This was no surprise to chemists, who had for some years been discovering new elements in nature and adding them to the table.

Mendeleev's periodic table of elements was nearly complete—only a few gaps remained. No trace in nature could be found, for example, of element 43, technetium, and element 61, promethium. At the time, the mystery could not be explained, as it was not yet known that the abundance of elements depended on their nuclear properties (unknown in those days) and not on their chemical properties.

Furthermore, the listing of the elements in ascending order of atomic weight seemed to terminate at element 92, uranium. There was no way of determining whether any elements existed beyond uranium, although the periodic system did not exclude their existence.

In subsequent decades, nuclear physics made it possible to clarify the concept of atomic number as the number of charges in the nucleus or the number of electrons contained in the atom, and also to clarify the concept of an isotope—a nucleus differing from another by the number of its neutrons but not by the amount of its charge. The concept of the isotope highlighted the reason for fractional values of atomic weights: an element as found in nature is a mixture of various isotopes, and even if each isotope were to have an integral atomic weight, the mixture of the isotopes—of practically constant composition throughout the crust of the Earth—must necessarily have a fractional value of mass. Subsequently, when accelerators and nuclear reactors permitted physicists to accelerate subnuclear particles or even whole nuclei, it became possible to fabricate new nuclei artificially. These new nuclei were isotopes of other elements that were already known or that showed the characteristics of elements that would occupy places beyond uranium in the periodic table.

ARTIFICIAL ISOTOPES AND TRANSURANIUM ELEMENTS

Nuclear physics states the conditions under which a nucleus is stable. If the protons and neutrons are to remain bound within the nucleus, it is essential that their unity reflect a ratio based on the number of protons. For light elements this ratio is of the order of 1; that is, there are as many protons as there are neutrons; in the case of the heavier elements, however, the neutron-proton ratio tends toward a value of 1.5. The neutrons, therefore, appear to be the binding agents needed to overcome the repulsive force of the protons, all of which carry positive charges and so would tend to repel each other. If the optimal number of neutrons is exceeded, the nucleus will become unstable—and likewise if the number of neutrons is insufficient. In the case of nuclei with high atomic numbers, however, the lack or the excess of one or even a few neutrons may still leave the nucleus stable.

If one of these stable isotopes is bombarded with a neutron in such a way that the neutron remains in the nucleus, the result is a nucleus with an excess neutron. Alternately, if the nucleus is bombarded with a proton, the result will be a nucleus with an excess proton. The same result can be obtained by bombarding the nucleus with a deuteron or an alpha particle, both of which have a proton-neutron ratio equal to 1. If the bombarded nucleus is a heavy one, it will acquire an excess of positive charge that will render it unstable.

Nuclei with an excess of protons are unstable and will emit positrons (positive electrons); those with an excess of neutrons will emit electrons. The radioactivity in the former case is thus $\beta+$, while in the latter case it is $\beta-$.

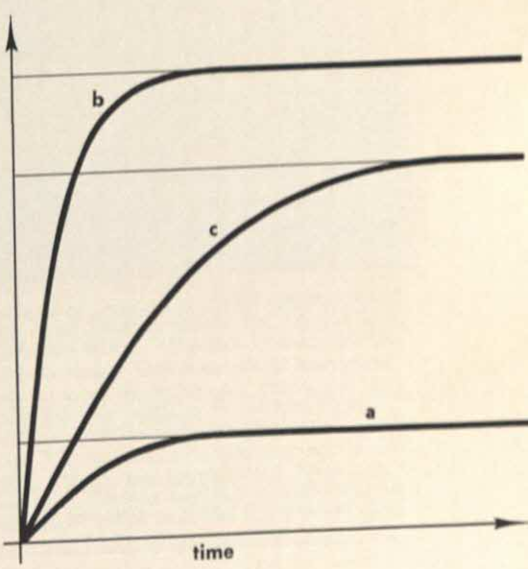
THE THEORY OF IRRADIATION—To produce an artificial radioisotope, a parent element must be bombarded with suitable particles. The parent may be stable or radioactive. If it is stable, or if it is radioactive with a long period of decay in relation to the length of the experiment, the bombardment of the parent substance can be carried out without concern for its decay during the experiment.

Even so, if a nucleus that does not exist in nature is produced, it will be radioactive and will, therefore, disintegrate during its production. If its disintegration is rapid and if a reasonable amount of the isotope is to be produced, the bombardment must be carried out with a beam of high intensity—a beam in which a relatively large number of particles passes per square centimeter per second. A weaker beam would result in nuclei disintegrating during the bombardment itself; at the end of the experiment the sample would be poorer in the parent substance and enriched in the nuclide of the second generation, while the daughter nuclide, the object of the experiment, would be very nearly absent.

Curve **a** in the graph shows how the number of daughter nuclides grows when the bombardment is effected by means of a low intensity beam and the daughter nuclide disintegrates rapidly. A saturation point is quickly reached, and the density of the daughter nuclide does not increase as the irradiation is prolonged. To obtain a greater density of the daughter element, a beam of higher intensity is necessary. Such a beam results in a greater saturation density, and the increased density of the daughter nuclide will be exactly proportional to the increase in the intensity of the bombarding beam, as shown by curve **b**. Beams of very high intensity are therefore used for the production of elements that decay very rapidly.

Curve **c**, on the other hand, shows the growth in the density of the daughter element when a beam of the same intensity as in **a** is used but when the daughter nuclide has a lower rate of decay. A higher density of the daughter element will now be attained, but the time required to reach this saturation density will be greater.

1



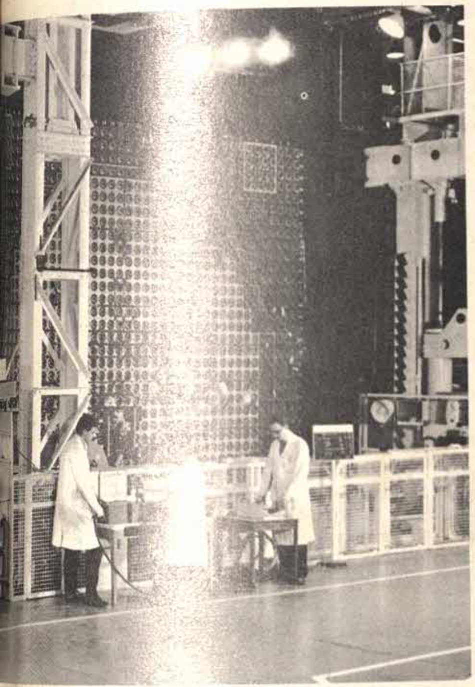
2

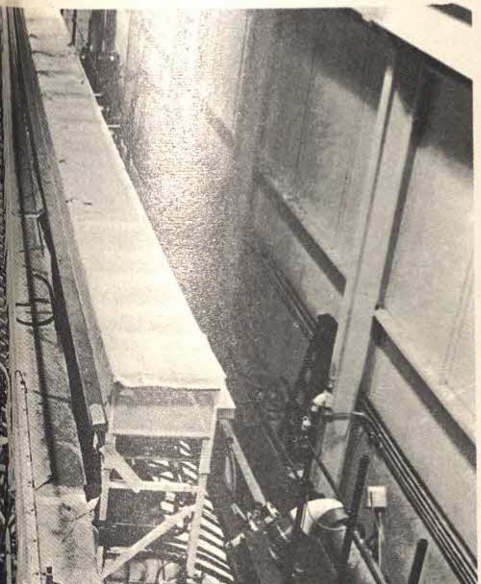
BEAM GENERATORS—In general, most artificial nuclides with atomic numbers below 92 are produced by means of nuclear reactors. The particle that carries out the bombardment is normally the neutron. In order to obtain high densities of daughter nuclides that decay very rapidly, it is essential to provide a very intense stream of neutrons.

For small bombardment experiments and the production of the more common radioisotopes, nuclear reactors such as that shown in Illustration 2a are used. Because such a reactor produces a moderately intense flow of neutrons in a fairly restricted space, its total power is very modest. A reactor of this type is easy to control by very simple methods. Such reactors are used for teaching purposes; they produce small quantities of artificial radioisotopes.

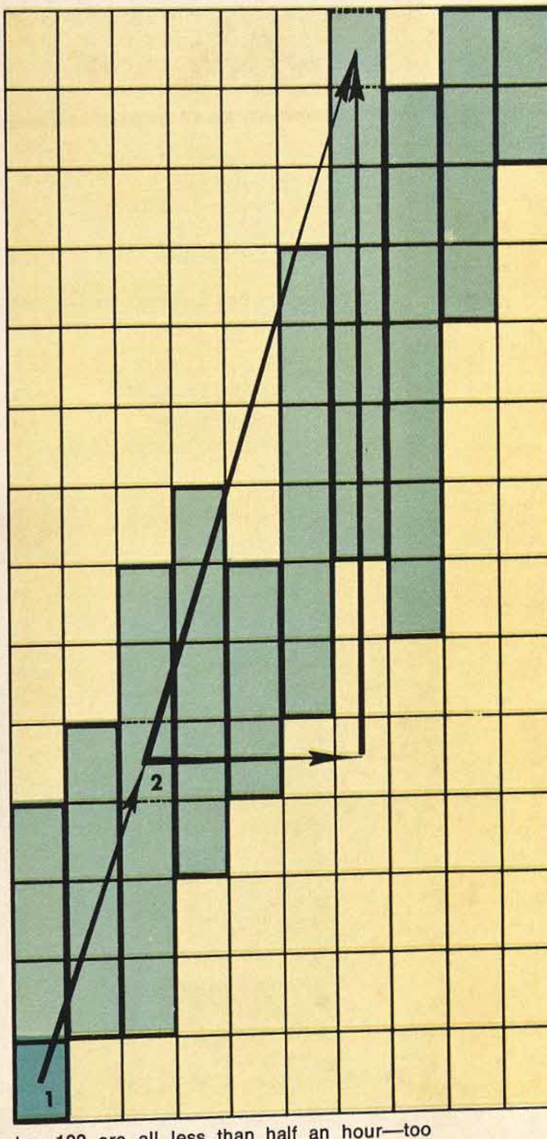
Where considerable quantities of radioactive materials with brief periods of decay are required, reactors that provide a very intense beam of neutrons are used. Such reactors (Illustration 2b) are often used to test the radioactivity of materials to be used in the construction of new types of reactors. To complete these tests in the shortest possible time, these testing reactors have very high neutron flows that will generate nuclides having very brief disintegration periods.

For the production of transuranium elements, beams of charged particles generated by an accelerator (Illustration 2c) are used.





Such particles are accelerated to energies between ½ Mev and 10 Mev. The beam of such an accelerator is concentrated on a well-defined point of the target by means of suitable focusing. This concentration makes it possible to obtain the high flux of charged particles required for producing nuclides at the required rate.



3

THE ASCENDING SCALE IN THE TABLE OF THE TRANSURANIUM ELEMENTS — This graph illustrates the principle followed in producing transuranium elements with a high atomic number. The graph represents a generic part of the table of the transuranium elements. The light green squares indicate the nuclides that have been produced. The dotted squares indicate the stable nuclides.

If the average life of an artificial nuclide is long—many days or years—the nuclide can be separated by chemical means after it has been produced by bombardment. This is a very useful operation because it concentrates the nuclide; therefore, its subsequent bombardment will be more efficient. In fact, it would be quite useless to have a very intense bombarding beam if the target contained only a few dispersed atoms of the parent element. Thus, a transuranium element with a high atomic number is produced from a parent element that is sufficiently stable and that can be concentrated by chemical means. Its neighbors are then examined in order to determine whether there exists another stable element two, three, or four atomic numbers further up the scale. Beginning with the parent element marked 1 on the graph, for example, a dotted square occurs two columns to the right and five columns up; it is, therefore, convenient to bombard the parent nuclide with alpha particles, which will increase its atomic number by two squares and its mass by four squares, thus bringing it exactly to the desired point 2. The next square with a sufficiently stable nucleus occurs four atomic numbers to the right and nine mass numbers further up. The nuclide best fitted to serve as a bombarding projectile is thus Be^9, which has an atomic number of 4 and will bring the nuclide exactly to the desired point. The explorative phase begins at this point, beyond which many concentrated targets of this last nuclide are prepared and bombarded with every available particle in order to determine if some particle will give rise to stable nuclides. It is unnecessary to isolate the different particles: the spectrum of the particles of disintegration is sufficient for ascertaining the presence of a new nuclide.

Attempts are being made to extend the periodic table of elements by bombarding nuclei with relatively heavy elements, such as carbon. Nevertheless, only nuclides with very rapid disintegrations are being obtained. The half-lives of nuclides above the atomic number 100 are all less than half an hour—too short a time to attempt chemical separation. It is probable that Mendeleev's periodic table cannot be extended much further, and that not many artificial nuclides will be produced beyond ^{256}Md and ^{254}No.

The very heavy elements are placed at the limit of the periodic table. If one of these elements is bombarded with heavy particles such as alpha particles, their mass number will increase by a value of 4 and their atomic number by a value of 2. If uranium (element 92) is bombarded, the end result is plutonium (element 94). This can also be produced by adding a neutron to uranium-238, as in a nuclear reactor. The element obtained this way can be bombarded a second time to obtain an element that is beyond uranium in the periodic table. Its mass number can be further increased by bombarding it once more. However, it is more practical to take longer steps and carry out bombardment with nuclei heavier than the helium nucleus—lithium and beryllium nuclei, for example. By such methods it has been possible to produce certain of the transuranium elements.

ABBREVIATIONS

A	ampere
Å	Ångstrom unit
abs	absolute
a-c	alternating current (as an adjective)
amu	atomic mass unit
atm	atmosphere
at. wt	atomic weight
AU	astronomical unit
avdp	avoirdupois
Bev	one billion electron volts
bhp	brake horsepower
bhp-hr	brake horsepower-hour
bp	boiling point
Btu	British thermal unit
C	temperature Celsius; temperature Centigrade
c	candle
cal	calorie
cfm	cubic feet per minute
cfs	cubic feet per second
cgs	centimeter-gram-second (system)
cl	centiliter
cm	centimeter
cm^2	square centimeter
cm^3	cubic centimeter
coef	coefficient
colog	cologarithm
cos	cosine
cot	cotangent
cp	candlepower
csc	cosecant
cu	cubic
cu ft	cubic foot
db	decibel
d-c	direct current (as an adjective)
doz	dozen
E	electromotive force
e	the base of the system of natural logarithms
ev	electron volt
F	temperature Fahrenheit
fp	freezing point
fpm	feet per minute
fps	feet per second
ft	foot; feet
ft^2	square foot
ft^3	cubic foot

ft-c	footcandle
ft-lb	foot-pound
G	universal gravitational constant
g	gram
gal	gallon
g-cal	gram-calorie
gpm	gallons per minute
gps	gallons per second
hr	hour
$h\nu$	photon energy
hp	horsepower
Hz	hertz (cycles per second)
I	electric current
ID	inside diameter
in.	inch
$in.^2$	square inch
$in.^3$	cubic inch
in.-lb	inch-pound
ips	inches per second
j	joule
K	temperature Kelvin (absolute)
kcal	kilocalorie
kg	kilogram
kg-cal	kilogram-calorie
kg-m	kilogram-meter
kg/m^3	kilograms per cubic meter
kgps	kilograms per second
km	kilometer
kv	kilovolt
kw	kilowatt
kwhr	kilowatt-hour
l	liter; lumen
lat	latitude
lb	pound
lb-ft	pound-foot
lb/ft^2	pounds per square foot
lb/ft^3	pounds per cubic foot
lb-in.	pound-inch
l-hr	lumen-hour
lin ft	linear foot
log	logarithm (common)
$\log_e$	logarithm (natural)
long.	longitude
m	meter; minute (time, in astronomical circles)

m^2	square meter
m^3	cubic meter
ma	milliampere
Mev	one million electron volts
mg	milligram
mh	millihenry
mi	mile
mi^2	square mile
min	minute
m-kg	meter-kilogram
ml	milliliter
mm	millimeter
mm^2	square millimeter
mm^3	cubic millimeter
$m\mu$	millimicron
mph	miles per hour
mphps	miles per hour per second
mv	millivolt
N	Avogadro constant
n!	factorial *n*
OD	outside diameter
oz	ounce
pH	rating on acid-alkaline scale
ppm	parts per million
psi	pounds per square inch
psia	pounds per square inch absolute
R	temperature Reaumur; resistance
RA	right ascension
rpm	revolutions per minute
rps	revolutions per second
sec	secant; second
sin	sine
sp gr	specific gravity
sq	square
tan	tangent
V	volt
VA	volt-ampere
W	watt; work
yd	yard
yd^2	square yard
yd^3	cubic yard

SCIENTIFIC SYMBOLS AND ABBREVIATIONS

α	alpha particle
β; β^-	beta particle
β^+	positron
γ	gamma radiation
Δ	a small change; heat
λ	wavelength; radioactive-decay constant
ma	milliampere
μc	microcurie
μf	microfarad
$\mu in.$	microinch
μm	micron
$\mu\mu$	micromicron
$\mu\mu f$	micromicrofarad
ν	frequency; neutrino
π	3.14159; osmotic pressure

Σ	the sum of
σ	nuclear cross section (barns); area
Ω	electrical resistance (ohms)
ω	angular speed; angular velocity
′	minute (angular measure)
″	second (angular measure)
♂	male
♀	female
>	is greater than
<	is less than
$\propto$	is proportional to
∞	infinity
$\sqrt{}$	square root of
°	degrees; temperature; angle measurement (example, 30°)

[ ]	molar concentration
+	positive electric charge; mixed with; plus
−	negative electric charge; single covalent bond; minus
=	equals; double covalent bond; produces
≠	does not equal
≡	triple covalent bond
→	produces; forms; chemical reaction
⇄	reversible chemical reaction
↑	gas produced by a chemical reaction
↓	precipitate produced by a chemical reaction
*	radioactive substance (follows symbol of element; example, Cl^*)

THE
ILLUSTRATED SCIENCE
DICTIONARY

Potential Difference to Reflux

KEY TO PRONUNCIATION

The diacritical marks are:

ə	b*a*nan*a*, *a*but
ᵊ	preceding l, m, n as in batt*le*
ə̇	*e*lectric
ər	f*ur*ther
a	m*a*t
ā	d*ay*
ä	c*o*t, f*a*ther
au̇	n*ow*, *ou*t
e	b*e*t
ē	b*ea*t
i	t*i*p
ī	b*i*te
j	*j*ob, *g*em
ŋ	si*ng*
ō	b*o*ne
ȯ	s*aw*, *a*ll
ȯi	c*oi*n
th	*th*in
$\underline{\text{th}}$	*th*en
ü	r*u*le, f*oo*l
u̇	p*u*ll, w*oo*d
ue	German h*ü*bsch
$\overline{\text{ue}}$	French r*ue*
yü	*u*nion
zh	vi*si*on

ˈ mark preceding the syllable with strongest stress.
ˌ mark preceding a syllable with secondary stress.

The system of indicating pronunciation in these volumes is used by permission from Webster's *Third New International Dictionary*, copyright 1961 by G. & C. Merriam Co., Publishers of the Merriam-Webster Dictionaries.

potential difference \pə-'ten-chəl 'dif-ərn(t)s\

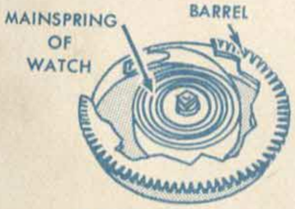
PHYSICS. A difference in the electrical condition of two points, causing electrons to flow from one point to the other when the points are connected by a good conductor of electricity. This difference is commonly measured in volts.

A new flashlight cell has a POTENTIAL DIFFERENCE *of 1.5 volts between its outer metal case and the central knob on its top.*

potential energy \pə-'ten-chəl 'en-ər-jē\

CHEMISTRY and PHYSICS. Stored mechanical or chemical energy that is able to do work because it is in a certain position or chemical condition. It is measured in such units as ergs, joules or foot-pounds.

Winding a watch places the spring under stress, giving it POTENTIAL ENERGY *that, when released, will run the watch.*



POTENTIAL ENERGY

potentiometer \pə-ˌten-chē-'äm-ət-ər\ *n.*

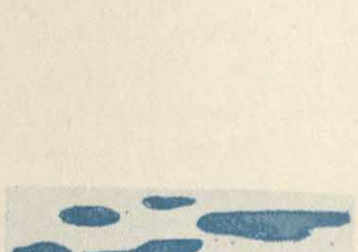
ENGINEERING and PHYSICS. A device used to measure an electromotive force or potential difference by comparing it with a known electromotive force or potential difference; also, a device used in electrical circuits to divide a given voltage into two parts; also called voltage divider.

A slide-wire POTENTIOMETER *is used in laboratories for precise measurement of voltages.*

pothole \'pät-ˌhōl\ *n.*

EARTH SCIENCE. A hole, usually deeper than it is wide, formed in rock by the grinding action of stones spun around by an eddy current.

The size of a POTHOLE *depends on the hardness of the rock and the duration of grinding action.*



power \'pau̇(-ə)r\ *n.*

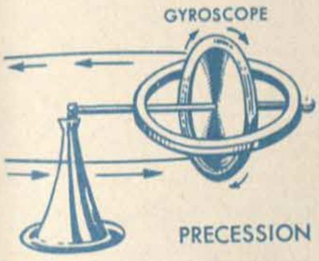
1. PHYSICS. The rate at which energy is used or at which work is done. This rate may be expressed in such units as foot-pounds per second, watts or horsepower and can be calculated by dividing the amount of work done by the time required to do the work. 2. MATHEMATICS. The product resulting from x multiplied by itself m times if m is a positive integer. It is written x^m. The ^{m} is called the exponent and x is called the base.

Since the first steam engines took over tasks formerly done by horses, the POWER *of engines has been expressed as the power of a certain number of horses.*

prairie \'pre(ə)r-ē\ *n.*

EARTH SCIENCE. An extensive tract of grassland on a plain. A prairie lacks large forests and is called a pampa in South America, a veld in Africa and a steppe in Eurasia.

A climate with low average rainfall and hot summer seasons contributes to the typical vegetation of a PRAIRIE.

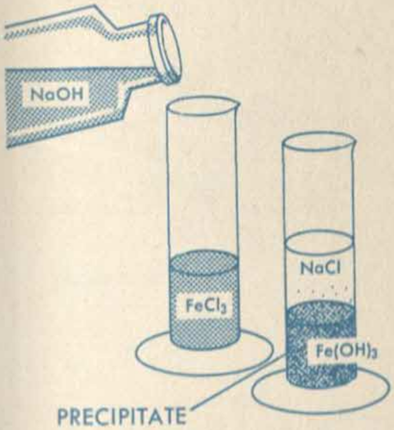


PRECESSION

precession \prē-'sesh-ən\ *n.*

1. PHYSICS. The circular motion of the upper end of the spin axis of a spinning top or other spinning, rigid body, resulting from torque applied to the body so as to change the direction of the axis of rotation. 2. ASTRONOMY. The slow change in the direction of the axis of rotation of the earth that results in a very slow westward motion of the equinoxes.

The rate of PRECESSION *of a spinning top or gyroscope increases as its spin axis is tipped farther away from a vertical line.*



PRECIPITATE

precipitate \pri-'sip-ət-ət\ *n.*

CHEMISTRY. Solid particles that form from a solution and that usually settle to the bottom of the solution.

The symbol ↓ after a chemical formula in a chemical equation means that the substance is a PRECIPITATE.

precipitation \pri-ˌsip-ə-'tā-shən\ *n.*

1. EARTH SCIENCE. Moisture that condenses out of the atmosphere and is deposited on the earth in such liquid or solid form as rain, hail, sleet, snow, dew and frost; generally, rainfall. 2. CHEMISTRY. The process or chemical reaction that results in the formation of a precipitate.

In measuring PRECIPITATION, *solid forms, such as hail and snow, are melted and recorded as inches of water.*

PREDATOR

predator \'pred-ət-ər\ *n.*

ZOOLOGY. Any animal that captures and eats other animals.

The fox is a PREDATOR *of rabbits.*

premolar \(')prē-'mō-lər\ *n.*

ANATOMY. Any one of the teeth situated just in front of the molars; also, a molar tooth that is normally shed and replaced during early childhood.

Because a PREMOLAR *has two points on its grinding surface, it is a bicuspid.*

preservative \pri-'zər-vət-iv\ *n.*

CHEMISTRY. A chemical substance added to food or other organic matter to prevent or delay spoiling or decay.

Sodium benzoate is often used as a PRESERVATIVE *for catsup.*

pressurize \'presh-ə-ˌrīz\ *v.*

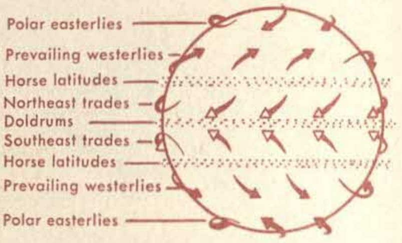
AERONAUTICS and ASTRONAUTICS. To produce and maintain a higher air pressure in an aircraft or space vehicle than the pressure of the surrounding environment.

It is necessary to PRESSURIZE *the suits of pilots and astronauts whose vehicles have less than atmospheric pressure.*

prevailing winds \pri-'vā(ə)l-iŋ 'windz\

EARTH SCIENCE. Winds at a given location that blow from a particular direction with greater frequency than from other directions.

The trade winds near latitude 30 degrees north are PREVAILING WINDS *blowing from the northeast.*



PREVAILING WINDS

primary colors \'prī-ˌmer-ē 'kəl-ərz\

PHYSICS. The three colors of light, orange-red, green and blue-violet, that can produce all other colors when mixed in various combinations.

When all three PRIMARY COLORS *are combined properly, the result can be near white.*

primary root \'prī-ˌmer-ē 'rüt\

BOTANY. The first root that develops from the seed of a plant.

A PRIMARY ROOT *that persists through the life of a plant is called a taproot.*

PRIMARY ROOT

primate \'prī-ˌmāt\ *n.*

ZOOLOGY. Any mammal belonging to the order of Primates, including man, the ape, the monkey, the marmoset and the lemur.

One distinguishing characteristic of a PRIMATE *is fingernails.*

prime factors \'prīm 'fak-tərz\

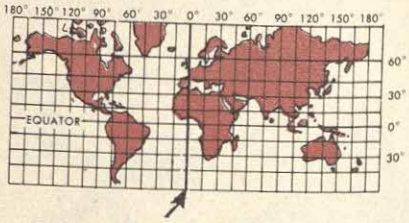
MATHEMATICS. The prime numbers or prime expressions whose product is a given number or expression.

The PRIME FACTORS *of 36 are 2, 2, 3 and 3.*

prime meridian \'prīm mə-'rid-ē-ən\

EARTH SCIENCE. An imaginary north-south line on the earth's surface. The line stands for 0 degrees longitude and divides the regions of east and west longitude. It passes through the Royal Observatory at Greenwich, near London, England.

Greenwich Mean Time is based on the position of the sun with relation to the PRIME MERIDIAN.



PRIME MERIDIAN

prime mover \'prīm 'mü-vər\

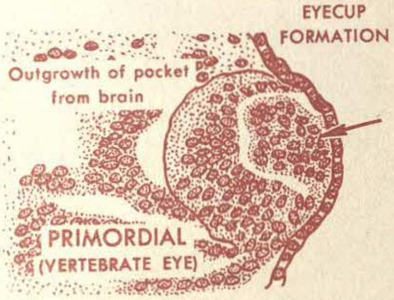
PHYSICS. Any engine that converts the energy of natural substances into mechanical energy.

A diesel engine is an efficient PRIME MOVER *that converts the chemical energy of oil and air into usable power.*

prime number \'prīm 'nəm-bər\

MATHEMATICS. A positive whole number that is divisible only by itself and by 1, as 2, 3, 5, 7, 11, 13 or 17.

Every PRIME NUMBER *except 2 is an odd number.*



PRIMORDIAL (VERTEBRATE EYE)

primordial \prī-'mȯrd-ē-əl\ *adj.*

1. BIOLOGY. Descriptive of the earliest recognizable form of a tissue or organ of an animal or plant. 2. Referring to substances existing at the beginning, such as the gases and dust that, according to some theories, were the origin of the stars, planets and satellites.

The PRIMORDIAL *germ cells of a male develop into sperm.*

principal root \'prin(t)-s(ə-)pəl 'rüt\

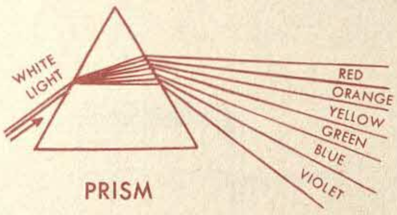
MATHEMATICS. With roots of positive numbers, the positive real root; with odd roots of negative numbers, the negative real root.

The PRINCIPAL ROOT *of a number depends on the index of the root.*

principle \'prin(t)-s(ə-)pəl\ *n.*

A general truth based on observed facts; also, an assumption derived from observed facts.

"The closer a planet is to the sun, the shorter its period of revolution," is a PRINCIPLE *of planetary motion.*



PRISM

prism \'priz-əm\ *n.*

1. PHYSICS. A solid having several flat surfaces, two of which are parallel polygons (bases), and having other surfaces in the shape of parallelograms. A prism is frequently made of glass with triangular bases. If transparent, it can form a spectrum by the refraction of white light. Some are used as high-efficiency mirrors in such optical devices as binoculars, where special techniques prevent the formation of a spectrum. 2. MATHEMATICS. A polyhedron having two faces, called bases, that are parallel, congruent polygons, and whose other faces are parallelograms.

A triangular PRISM *of clear rock salt can form a spectrum from infrared radiation as well as a visible spectrum from light.*

probability \ˌpräb-ə-ˈbil-ət-ē\ *n.*

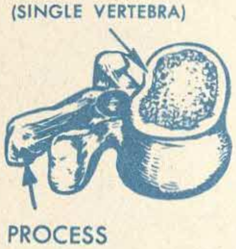
MATHEMATICS. The likelihood, or chance, that a given event or group of events will occur in a specified way; the ratio m/(m+n), where m equals all the ways a given event may succeed or occur and n equals all the ways it may fail or not occur.

The PROBABILITY *that a 4 will turn up on one throw of a die is one in six, or* $\frac{1}{6}$.

proboscis \prə-ˈbäs-əs\ *n.*

ZOOLOGY. Any tubular elongation, usually of the head; in mammals, a long, flexible snout, as of a tapir or an elephant; in insects, the extended mouth parts; in planaria, the feeding tube.

The PROBOSCIS *of an elephant is used in feeling, smelling, pushing, drinking, eating, bathing and breathing.*



process \ˈpräs-ˌes\ *n.*

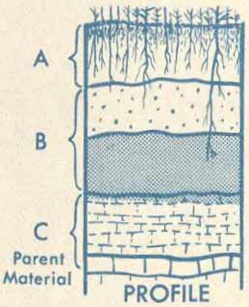
1. ANATOMY. An elongated or projecting part of a larger structure, as any one of the projecting parts found on the bones of the vertebral column. 2. A method of doing or making something, or the manner in which something occurs, usually made up of several to many steps.

In man and most other vertebrate animals, each rib fits into an articular PROCESS *of a vertebra, permitting a slight movement of the rib during breathing.*

product \ˈpräd-(ˌ)əkt\ *n.*

1. MATHEMATICS. The answer that results when one number or quantity is multiplied by another. 2. CHEMISTRY. The substance or substances resulting from a chemical reaction. A product has different properties from those of the material or materials that reacted.

The PRODUCT *of .5 and 7 equals 3.5.*



profile \ˈprō-ˌfīl\ *n.*

EARTH SCIENCE. A vertical section showing the layers making up a soil; also, a drawing showing an earth surface outline and the vertical arrangement of materials beneath the surface.

The different layers in a soil PROFILE *have different physical and chemical characteristics.*

profundal \prə-ˈfənd-ᵊl\ *adj.*

BIOLOGY and EARTH SCIENCE. Pertaining to the deep-water area of freshwater bodies, usually limited to areas free of rooted plants.

Common freshwater clams are residents of the PROFUNDAL *zone.*

program \'prō-ˌgram\
ENGINEERING and MATHEMATICS (N.). The sequence of steps, or operations, performed by a computing machine in solving a problem. (V.). To work out a sequence of arithmetic operations to solve a given type of mathematical problem with a computer.

If the PROGRAM *for a computer is stored in a memory unit, the calculation may be performed many times.*

WORLD MAP

PROJECTION
(AITOFF)

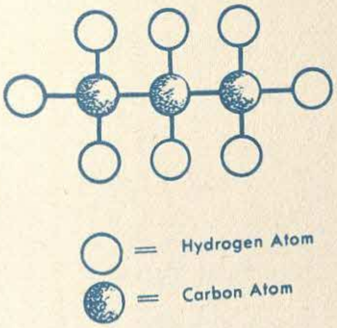
projection \prə-'jek-shən\ *n.*
MATHEMATICS. A mapping of points and lines of one plane in some specified manner upon another plane; more generally, any mapping of the points of a given surface upon another surface, as in cartography and photography; also, the image produced by such a mapping.

The plane on which a figure is projected is called the plane of PROJECTION.

propagation \ˌpräp-ə-'gā-shən\ *n.*
1. BOTANY. Multiplication of plants by natural or artificial separation of their vegetative parts from the parent plant. The vegetative parts root and develop as individual plants. 2. PHYSICS. A process by which such energy as sound or light is transmitted through a substance. 3. CHEMISTRY. The process by which a flame travels through a burnable mixture of gases.

Root cutting is an artificial method of PROPAGATION *used in the cultivation of sweet potatoes.*

PROPANE
(MOLECULAR STRUCTURE)



propane \'prō-ˌpān\ *n.*
CHEMISTRY. A colorless, flammable gas at atmospheric pressure; a hydrocarbon C_3H_8, that may be liquefied under pressure.

PROPANE *is compressed in steel cylinders and sold as a fuel for home and industrial use.*

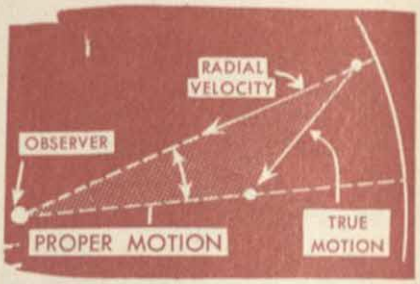
proper fraction \'präp-ər 'frak-shən\
MATHEMATICS. In arithmetic, a fraction whose value is greater than zero and less than one or whose numerator is less than the denominator; in algebra, a fraction whose numerator is of lower degree than the denominator for the variable or variables appearing in them.

The expression $\frac{7}{16}$ *is a* PROPER FRACTION.

proper motion \'präp-ər 'mō-shən\

ASTRONOMY. The angular rate of change in the position of a star on the celestial sphere.

PROPER MOTION is measured by the relative positions of a star on photographs taken over a period of years.



prophase \'prō-ˌfāz\ *n.*

BIOLOGY. A stage during mitotic cell division, characterized by the appearance of distinct chromosomes, the formation of a spindle and the dissolving of the nuclear membrane.

PROPHASE is preceded by interphase and followed by metaphase.

proportion \p(r)ə-'pōr-shən\ *n.*

MATHEMATICS. A statement of the equality of two ratios.

$3{:}6 = 5{:}10$ is a PROPORTION.

protective coloration \prə-'tek-tiv ˌkəl-ə-'rā-shən\

BIOLOGY. Color patterns or markings that blend with color patterns in an organism's usual environment and that seem to camouflage and protect the organism.

The PROTECTIVE COLORATION of the horned toad consists of markings in shades of brown that blend with desert colors.

PROTECTIVE COLORATION

protein \'prō-ˌtēn\ *n.*

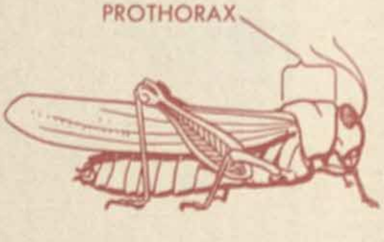
CHEMISTRY. Any one of a large group of complex organic compounds composed of many amino acid molecules joined to make one very large molecule. Proteins contain the elements carbon, hydrogen, nitrogen, oxygen and frequently sulfur and phosphorus, and are essential to all known living organisms.

Concentrated nitric acid applied to a substance containing a PROTEIN produces a yellow color.

prothorax \(')prō-'thō(ə)r-ˌaks\ *n.*

ZOOLOGY. The first, or front, segment of the three segments in the thorax of insects. The prothorax bears the first pair of legs.

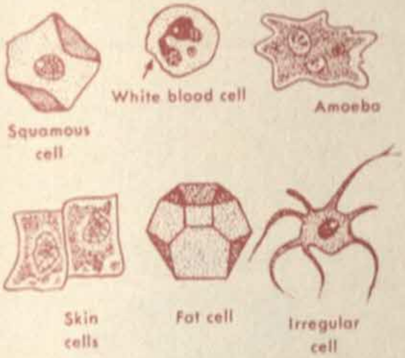
The PROTHORAX of the locust includes a saddlelike covering usually extending back to the base of the wings.



proton \'prō-ˌtän\ *n.*

CHEMISTRY and PHYSICS. A subatomic particle with an atomic mass of one (about 1,836 times the mass of an electron) and an electric charge of plus one. It forms the nucleus of an ordinary hydrogen atom, and at least one is contained in the nucleus of every atom.

If a PROTON *is injected into an atomic nucleus, the atomic number of the nucleus is increased by one.*



PROTOPLASM

protoplasm \'prōt-ə-ˌplaz-əm\ *n.*

BIOLOGY. The complex matter of which living things are composed; living matter; living material; living substance.

Although many of the properties and chemical substances of PROTOPLASM *are known, it has not yet been synthesized in a laboratory.*

protoplast \'prōt-ə-ˌplast\ *n.*

BOTANY. The protoplasm of a single plant cell.

A cell wall encloses, but is not part of, the PROTOPLAST.

prototype \'prōt-ə-ˌtīp\ *n.*

1. BIOLOGY. The original ancestral form of a living organism.
2. ENGINEERING. The original model of a design or product on which later models are based.

The PROTOTYPE *of a whale was a land animal with four legs.*

PROTOZOAN
PARAMECIUM

protozoan \ˌprōt-ə-'zō-ən\ *n.*

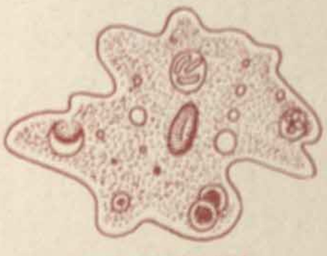
ZOOLOGY. Any animal belonging to the phylum Protozoa, which includes one-celled animals that are usually microscopic and water-dwelling.

A parasitic PROTOZOAN *may be found in the body of a larger animal.*

proximal \'präk-sə-məl\ *adj.*

ANATOMY and BIOLOGY. Referring to the area near a central point or near the attachment point of a limb or other part of an animal or organism, as opposed to distal.

A PROXIMAL *fracture of the thighbone is frequently set with a steel pin.*



PSEUDOPODIUM
AMOEBA

pseudopodium \ˌsüd-ə-'pōd-ē-əm\ *n.*

ZOOLOGY. In amoebas and similar kinds of one-celled organisms, a temporary flowing extension of protoplasm that is used in locomotion and ingestion.

White blood corpuscles pass through capillary walls by extending a PSEUDOPODIUM *between the cells of the vessel wall.*

P-shell \'pē-ˌshel\ *n.*

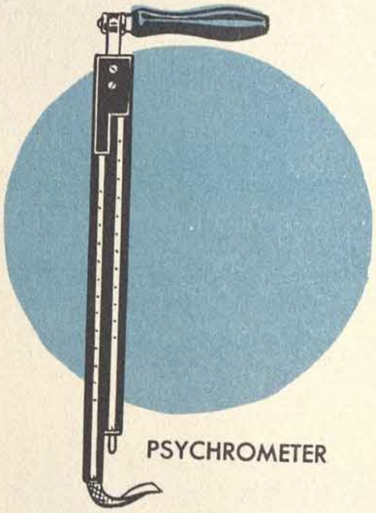
CHEMISTRY and PHYSICS. One of the seven principal energy levels, identified by the letters K, L, M, N, O, P and Q, that electrons may occupy in an atom; the sixth principal energy level that electrons may occupy in the Bohr model for the atom.

An electron in the P-SHELL *may lose energy and drop to an inner shell that is not full of electrons.*

psittacosis \\ˌsit-ə-ˈkō-səs\\ *n.*

MEDICINE and ZOOLOGY. An infectious disease of birds that can be transmitted to humans. In birds, the disease is marked by diarrhea and dehydration, but in man it takes a form similar to pneumonia.

Imported birds are quarantined to insure against their having PSITTACOSIS.



PSYCHROMETER

psychedelic \\ˌsī-kə-ˈdel-ik\\ *adj.*

MEDICINE. Referring to a calm and pleasant mental state in which the senses are attuned to beauty and one feels capable of great creative effort; also, referring to drugs that are supposed to induce this state but which may also induce erratic or dangerous behavior and symptoms of severe mental illness.

LSD, psilocybin and mescaline are PSYCHEDELIC *drugs.*

psychiatry \\sə-ˈkī-ə-trē\\ *n.*

MEDICINE. The branch of medicine that deals with the diagnosis and treatment of mental illnesses.

Methods of PSYCHIATRY *have made it possible to cure a number of serious mental disorders.*

psychrometer \\sī-ˈkräm-ət-ər\\ *n.*

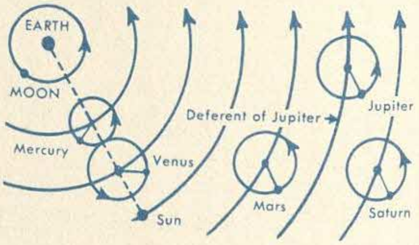
ENGINEERING and PHYSICS. A device used to measure the relative humidity of the air. It is usually constructed of two thermometers, one of which has its bulb covered with a wet cloth, that are mounted on a frame and whirled in the air.

A PSYCHROMETER *is often used in factories manufacturing paper, where the control of relative humidity is important.*

Ptolemaic system \\ˌtäl-ə-ˈmā-ik ˈsis-təm\\

ASTRONOMY. An explanation of the universe made in the second century A.D. by an Egyptian astronomer, Ptolemy, who believed that the earth was the center of the universe and that all the celestial bodies revolved around it on deferents and epicycles; see *Copernican system.*

The PTOLEMAIC SYSTEM *was challenged in the sixteenth century by the theories of Copernicus.*

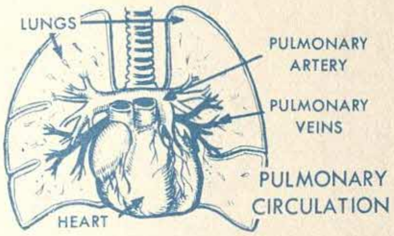


PTOLEMAIC SYSTEM

puberty \\ˈpyü-bərt-ē\\ *n.*

PHYSIOLOGY. A period occurring between childhood and adolescence in both sexes, during which physical and endocrine changes take place, equipping the body for adulthood and parenthood.

In temperate zones, PUBERTY *occurs in boys at the approximate age of 14 to 16 years, while in girls it occurs at an age of about 12 to 14 years.*



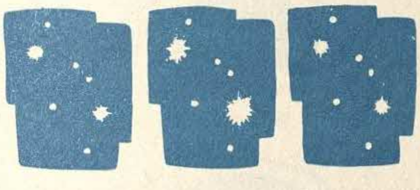
PULMONARY CIRCULATION

pulmonary circulation \'pu̇l-mə-ˌner-ē sər-kyə-'lā-shən\
PHYSIOLOGY. The flow of blood from the right ventricle of the heart through the pulmonary arteries into the capillaries of the lungs and back to the left side of the heart through the pulmonary veins.

In PULMONARY CIRCULATION, *blood going to the lungs is dark red, while aerated blood leaving the lungs is bright red.*

pulsating stars \'pəl-ˌsāt-iŋ 'stärz\
ASTRONOMY. A class of variable stars having regular periods of expansion and contraction that result in variations in brightness.

PULSATING STARS *have periods of variable brightness that range from a few hours to more than 300 days.*



PULSATING STARS

pulse \'pəls\ *n.*
1. PHYSIOLOGY. A rhythmic expansion of the arteries resulting from the contractions of the heart muscles and from the accompanying increase in blood pressure. 2. PHYSICS. An electrical signal, a sound signal or a force of short duration; also, a single wave.

Anger, fear and violent exercise will all increase the rate of the PULSE.

pulse-jet engine \'pəls-ˌjet 'en-jən\
AERONAUTICS. A type of jet engine that takes in air through inlet valves, which then close. The air is mixed with fuel and ignited. Expansion of gases provides thrust and creates a low pressure near the inlet valves, reopening them. The cycle is repeated several hundred times each minute.

The PULSE-JET ENGINE, *like other types of jet engines, is limited to regions of the atmosphere that contain enough oxygen to oxidize the fuel.*

PULSE

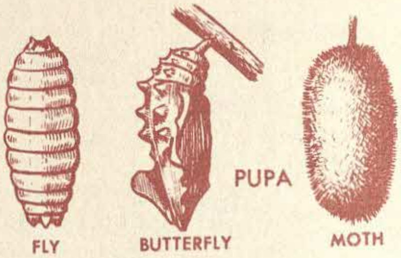
pumice \'pə-məs\ *n.*
EARTH SCIENCE. A porous, volcanic glass containing many small bubbles or cells. It is usually white or gray and occurs in layers or heaps around a volcanic vent. Pumice is used as an abrasive and as an aggregate in lightweight concrete blocks.

Because it contains many bubbles, PUMICE *has a specific gravity less than one and will therefore float in water.*

pupa \'pyü-pə\ *n.*

ZOOLOGY. The form of an insect undergoing the third of the four stages of metamorphosis. The form in this stage shows little activity as it changes from larva to adult. It is most often enclosed within a cocoon or other protective enclosure.

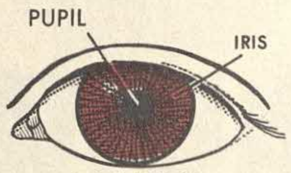
Although the mosquito PUPA *is very active, it does not feed until it emerges from the water as an adult.*



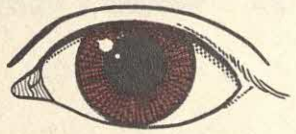
pupil \'pyü-pəl\ *n.*

ANATOMY and ZOOLOGY. In vertebrates and a few other animals, the circular or oval-shaped opening surrounded by the iris of the eye. It permits light to pass into the back part of the eyeball.

In man, the size of the PUPIL *decreases as the intensity of light striking the eye increases.*



In bright light



In dim light

purebred \'pyu̇(ə)r-'bred\ *adj.*

ZOOLOGY. Referring to animals with a large number of common hereditary characteristics that have been developed by inbreeding.

PUREBRED *animals may possess and pass on undesirable characteristics that increase in frequency through successive generations.*

putrefaction \ˌpyü-trə-'fak-shən\ *n.*

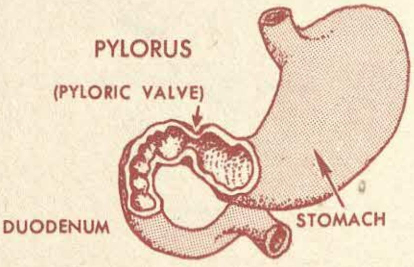
BIOLOGY. The gradual decay of living or dead organic matter, caused by the action of microorganisms.

PUTREFACTION *of dead plants and animals provides nitrates and other materials necessary to plant growth.*

pylorus \pī-'lōr-əs\ *n.*

ANATOMY. The opening between the stomach and the small intestine. Its contracting and relaxing muscle bands determine the rate at which the stomach empties food into the small intestine.

Contractions of muscles surrounding the PYLORUS *keep the stomach from forcing its entire contents into the small intestine at one time.*



pyramid \'pir-ə-ˌmid\ *n.*

MATHEMATICS. A polyhedron with four or more faces. One face, called the base, is a polygon that may have any number of sides, while the other faces, called lateral faces, are all triangles having one point, called the vertex of the pyramid, in common.

The altitude of a PYRAMID *is a perpendicular extending from the vertex of the pyramid to the plane of the base.*

pyrenoid \pī-'rē-ˌnȯid\ *n.*

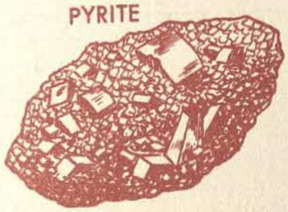
BOTANY. A colorless protein body in the pigment cells of certain algae and liverworts. It is a repository for starch.

The young cells of the algae Hydrodictyon each contain one PYRENOID, *while older cells contain many.*

pyrite \'pī(ə)r-ˌīt\ *n.*

CHEMISTRY and EARTH SCIENCE. Any of a group of metallic iron sulfide minerals commonly found with quartz, coal, copper and other minerals. Its composition is usually FeS_2.

Iron PYRITE, *also known as fool's gold, resembles small specks of gold embedded in rock.*



PYRITE

pyrolysis \pī-'räl-ə-səs\ *n.*

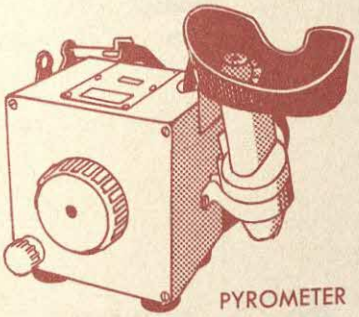
CHEMISTRY. The chemical change in a substance when heat is applied. The product of pyrolysis results from a rearrangement of atoms.

The PYROLYSIS *of red mercuric oxide produces oxygen and mercury.*

pyrometer \pī-'räm-ət-ər\ *n.*

ENGINEERING. An instrument used to measure very high temperatures. It may operate on electrical principles, such as the thermoelectric effect, or it may utilize the principle that the electrical resistance of platinum increases with increasing temperatures.

For measuring extremely-high temperatures, an optical PYROMETER, *utilizing the principle that the upper limit of the spectrum of an incadescent substance shifts to blue as the temperature increases, is used.*



PYROMETER

pyrophoric fuel \ˌpī-rə-'fȯr-ik 'fyü(-ə)l\

ASTRONAUTICS. A rocket propellant that is self-igniting when it comes in contact with oxygen.

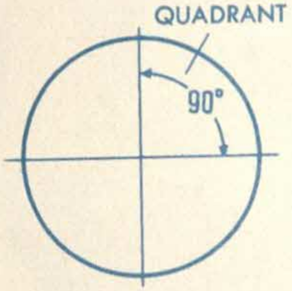
Hydrazine is a PYROPHORIC FUEL *that contains hydrogen and nitrogen.*

Pythagorean theorem \pə-ˌthag-ə-'rē-ən 'thē-ə-rəm\

MATHEMATICS. A proposition stating that the square on the hypotenuse of a right triangle is equal to the sum of the squares on the other two sides.

Surveyors often measure distances indirectly by methods based on the PYTHAGOREAN THEOREM.

Q



Q-shell \'kyü-ˌshel\ *n.*

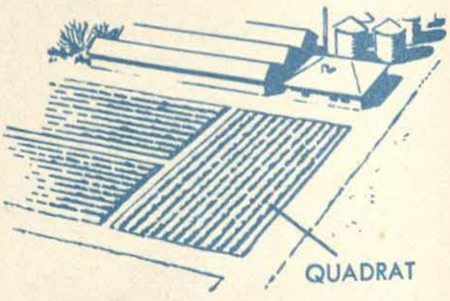
CHEMISTRY and PHYSICS. One of the seven principal energy levels, identified by the letters K, L, M, N, O, P and Q, that electrons may occupy in an atom; the seventh, or highest, principal energy level that electrons may occupy in the Bohr model for the atom.

In the periodic table of elements, the Q-SHELL *appears first in the element francium.*

quadrant \'kwäd-rənt\ *n.*

1. MATHEMATICS. One of the four parts created in a plane by two perpendicular lines; also, a quarter of a circle, or an arc equal to $\pi/2$ radians, or 90 degrees. 2. ENGINEERING. An instrument used in measuring altitudes.

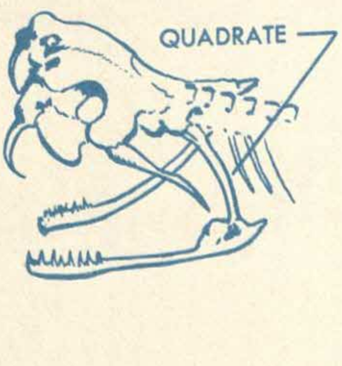
The first QUADRANT *of a rectangular coordinate plane is the portion lying above the x-axis to the right of the y-axis.*



quadrat \'kwäd-rət\ *n.*

BIOLOGY and EARTH SCIENCE. A rectangular sample plot of land used in taking a census of plants and animals that live in an area; also, a plot of land used for experimental purposes.

The study of a QUADRAT *on the north slope of a hill usually yields different results from a similar investigation of the south slope.*



quadrate \'kwäd-ˌrāt\ *n.*

ZOOLOGY. In most vertebrates below mammals, a bone or cartilaginous structure that joins the lower jaw to the skull. It is especially prominent in reptiles.

The QUADRATE *has evolved into a bone of the inner ear in mammals.*

quadratic equation \kwä-'drat-ik i-'kwā-zhən\

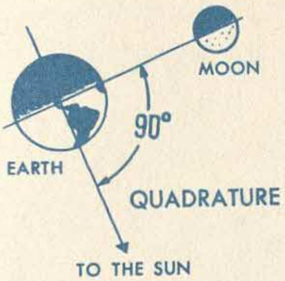
MATHEMATICS. An equation, or statement of equality between two expressions, in which 2 is the highest power of the unknown.

An equation such as $4x^2 + 2x + 6 = 0$ *is a* QUADRATIC EQUATION.

quadratic formula \kwä-'drat-ik 'fȯr-myə-lə\

MATHEMATICS. The formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, in which values for a, b and c in an equation of the form $ax^2 + bx + c = 0$ are substituted in order to find its roots.

The QUADRATIC FORMULA *can be used to find that the roots of* $2x^2 + 3x - 5 = 0$ *are* $x = 1$ *and* $x = -5/2$.



quadrature \'kwäd-rə-ˌchu̇(ə)r\ *n.*

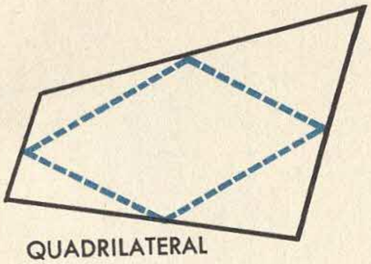
1. ASTRONOMY. The relationship of two heavenly bodies when lines drawn from each of them to the center of the earth form a 90-degree angle. 2. MATHEMATICS. The process of squaring, or determining the dimensions of a square whose area is equal to that of another given surface.

When the moon is in QUADRATURE *with the sun, half of the moon appears illuminated, and it is at meridian either at sunrise or sunset.*

quadrilateral \ˌkwäd-rə-'lat-ə-rəl\ *n.*

MATHEMATICS. A polygon having four sides.

The lines joining the midpoints of any QUADRILATERAL *form a parallelogram.*


QUADRILATERAL

quadruped \'kwäd-rə-ˌped\ *n.*

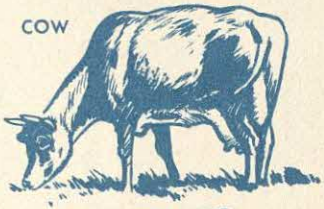
ZOOLOGY. A four-footed animal.

The cow is a QUADRUPED.

qualitative \'kwäl-ə-ˌtāt-iv\ *adj.*

Referring to the nature or kind of characteristics, properties or components, as in the chemical analysis of a substance wherein the kinds of materials are determined but not the amounts; see *quantitative.*

The QUALITATIVE *analysis of a sample of gasoline would indicate what compounds are present but not the amount of each.*



QUADRUPED

quantitative \'kwän(t)-ə-ˌtāt-iv\ *adj.*

Referring to the properties of substances or processes, such as their mass, length, speed or duration, as in the analysis of a substance to determine how much of each component exists; see *qualitative.*

One QUANTITATIVE *difference between two objects can be determined by comparing their weights.*

quantum \'kwänt-əm\ *n.*

PHYSICS. One of the separate, or discrete, quantities of energy set forth in the quantum theory of radiation.

The formula for the energy of a QUANTUM *is* $E = h\nu$, *in which* ν *equals frequency of energy involved and* h *equals Planck's constant* (6.55×10^{-27}).

quantum mechanics \'kwänt-əm mi-'kan-iks\

MATHEMATICS and PHYSICS. Mathematical or mechanical applications of the quantum theory of atomic structure and interaction of matter and radiation.

QUANTUM MECHANICS *involves only observable and measurable phenomena, such as the frequency and intensity of spectral lines.*

QUARRY

quantum theory of radiation \'kwänt-əm 'thē-ə-rē əv ˌrād-ē-'ā-shən\

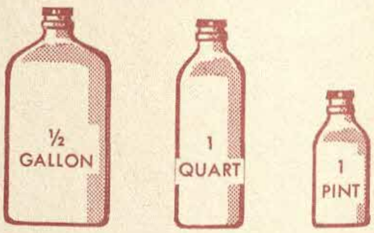
PHYSICS. An explanation for effects associated with light and other electromagnetic radiation that depicts the energy as occurring in bundles, or packets, of energy (photons). The theory has been used to explain emission and absorption of electromagnetic radiation.

The QUANTUM THEORY OF RADIATION *includes the wave theory of radiation in explaining phenomena such as refraction and interference.*

quarantine \'kwȯr-ən-ˌtēn\ *n.*

MEDICINE. The period of isolation of persons or animals coming from an area where a contagious disease is known to be, or suspected of being, present; also, prohibiting entrance into an area of plants and animals from a region known to be infested with disease.

The QUARANTINE *placed on cattle from some countries prevents the spread of such infections as hoof-and-mouth disease.*



QUART

quarry \'kwȯr-ē\ *n.*

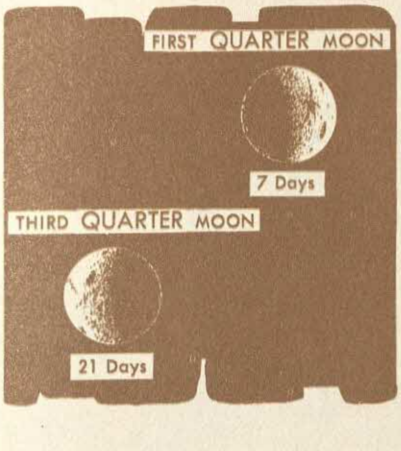
EARTH SCIENCE. An open excavation or pit from which stone, such as granite, slate or limestone, is extracted, often for use in the construction of buildings or roads.

An excavation made in searching for fossils is a QUARRY *of special interest to the paleontologist.*

quart \'kwȯ(ə)rt\ *n.*

MATHEMATICS. A standard unit of volume for both dry and liquid quantities; in dry measure, equal to about 67.2 cubic inches, or 1.101 liters; in liquid measure, equal to 57.75 cubic inches, or .946 liters; also, ⅛ peck or 2 pints; *abbr.* qt.

Milk is sold by the QUART.



quarter \'kwō(r)t-ər\ *n.*
1. ASTRONOMY. One of the four divisions of a revolution of the moon. 2. MATHEMATICS. One of the four equal parts into which something has been divided, or ¼ quantity.

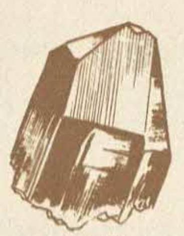
During first QUARTER, *the moon appears as a crescent.*

quartz \'kwȯ(ə)rts\ *n.*
CHEMISTRY and EARTH SCIENCE. SiO_2. A common mineral, composed of silicon and oxygen, that is clear, milky or colored. It is found in crystalline and apparently uncrystallized varieties and has a hardness of 7 on Mohs' scale.

QUARTZ *is one of the most important rock-forming minerals.*

quasar \'kwā-zər\ *n.*
ASTRONOMY. A quasi-stellar radio source; any extremely powerful radio source from 450 to 4,000 million light-years away that cannot be identified as a star or a galaxy.

The furthest-known object from the earth is the QUASAR *3C 147.*



QUARTZ

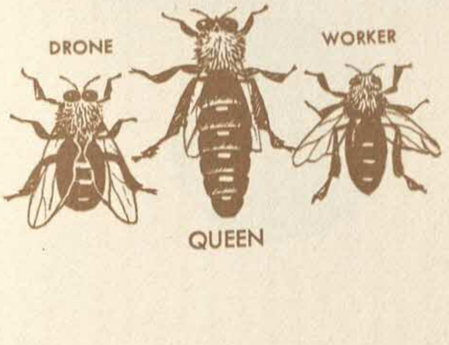
queen \'kwēn\ *n.*
ZOOLOGY. The reproductive female, or mother, of an insect society, as of bees, ants or termites. The queen is usually of large size and distinctive structure.

A larval female honeybee will be a worker if fed honey or a QUEEN *if fed royal jelly throughout development.*

quicksilver \'kwik-ˌsil-vər\ *n.*
Another name for the element mercury. See *mercury.*

quinary \'kwī-nə-rē\ *adj.*
MATHEMATICS. Referring to fives, arranged by fives or using five as a base, as the quinary system of numeration.

In the QUINARY *system of numeration, the number 37 (base 10) is represented by 122 (one 25, two fives and two units).*



quinine \'kwī-ˌnīn\ *n.*
CHEMISTRY and MEDICINE. $C_{20}H_{24}N_2O_2$. A chemical substance having alkaline properties. It is a diacid base and is obtained from the bark of cinchona trees that grow in South America.

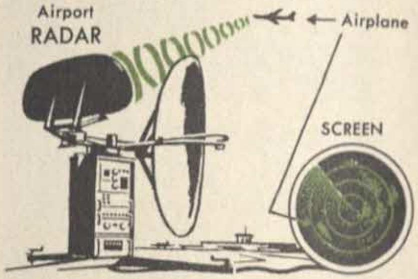
Salts of QUININE *are especially useful in treating malaria.*

R

radar \'rā-ˌdär\ *n.*

ENGINEERING and PHYSICS. An electronic system that sends radio waves and receives and analyzes their reflected waves, or echoes. It is used to determine the direction of, and distance to, an object, and uses radio wavelengths in the range of 0.75 cm. to 1.0 meter.

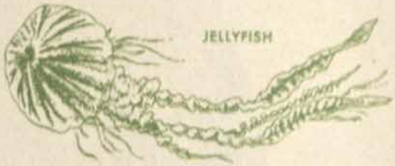
RADAR is used by some weather stations to locate cloud masses, hurricanes and other storm activity.



radial symmetry \'rād-ē-əl 'sim-ə-trē\

ZOOLOGY. The form of an animal body in the shape of a cylinder or bowl, with a central axis around which similar body parts are arranged, as in jellyfish and corals; see *biradial symmetry.*

An animal with RADIAL SYMMETRY can be cut into two identical halves by a lengthwise cut in any direction through the central axis.



RADIAL SYMMETRY

radial velocity \'rād-ē-əl və-'läs-ət-ē\

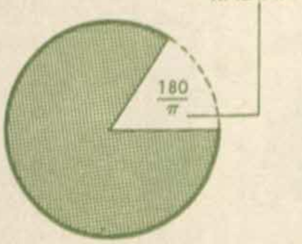
ASTRONOMY. The velocity, in an observer's line of sight, at which a star is approaching or going away.

A spectroscope can be used to determine the RADIAL VELOCITY of some stars.

radian \'rād-ē-ən\ *n.*

MATHEMATICS. A unit of angular and circular measure based on the radius of a circle. It is an angle that, as a central angle of a circle, intercepts an arc equal to the radius of that circle, and, therefore, is an arc of a circle equal in length to the radius of that circle.

Because the circumference of a circle is 2 π times its radius, 2 π radius equal 360 degrees, π radians equal 180 degrees and one RADIAN equals 180 degrees divided by π, or about 57.3 degrees.



radiant energy \'rād-ē-ənt 'en-ər-jē\

PHYSICS. Energy that travels as electromagnetic waves, including such energy as light, X rays and radio waves.

All RADIANT ENERGY travels at the speed of light.

radiation \\,rād-ē-'ā-shən\\ *n.*

PHYSICS. The process by which energy, such as subatomic particles, is emitted and transmitted. The process is characterized by transmission in straight lines from a source through a given medium, such as air, water or a vacuum; also, the transmitted energy itself.

RADIATION *of X rays occurs when high-speed electrons strike a metal target.*

radiation chemistry \\,rād-ē-'a-shən 'kem-ə-strē\\

CHEMISTRY. The study of chemical reactions caused or started by the absorption of alpha rays, gamma rays or other forms of radiant energy.

Scientists working in RADIATION CHEMISTRY *hope to understand better the effects of ionizing radiation on living organisms.*

RADIATION FOG

radiation fog \\,rād-ē-'ā-shən 'fȯg\\

EARTH SCIENCE. Condensation of water vapor in the air near the earth's surface because of cooling by radiation; ground fog.

RADIATION FOG *forms on calm, clear nights and disappears when heat from the sun raises air temperature above the dew point.*

radiation pressure \\,rād-ē-'ā-shən 'presh-ər\\

ASTRONOMY and PHYSICS. The force exerted by radiation on a unit area of an object.

Although RADIATION PRESSURE *is slight, it has a marked effect on such particles as gas ions and electrons.*

COMMON RADICALS

$C_2H_3O_2$ (Acetate)
NH_4 (Ammonium)
AsO_4 (Arsenate)
HCO_3 (Bicarbonate)
HSO_4 (Bisulfate)
CO_3 (Carbonate)
ClO_3 (Chlorate)
OH (Hydroxide)
NO_3 (Nitrate)
PO_4 (Phosphate)
SO_4 (Sulfate)
SO_3 (Sulfite)

RADICAL

radical \\'rad-i-kəl\\ *n.*

1. CHEMISTRY. A group of atoms that acts as a single unit in chemical reactions and in the formation of certain compounds. A radical may act as an ion, as do NH_4^+ and SO_4^{--}, or as a unit that can form a covalent bond, as do the methyl group, CH_3, and the benzyl group, $C_6H_5CH_2$. 2. MATHEMATICS. An indicated root of a number; also, the expression used to indicate the principal root of a number or quantity.

The ammonium RADICAL, NH_4^+, *and the nitrate radical,* NO_3^-, *are combined in equal numbers to form the crystalline compound ammonium nitrate.*

radical sign \\'rad-i-kəl 'sīn\\

MATHEMATICS. The symbol $\sqrt{}$ used to indicate that a root, shown by the index number, as $\sqrt[3]{}$, of a number or quantity is to be extracted.

The RADICAL SIGN *used without an index number indicates that the principal square root of the radicand is to be found.*

radicand \ˌrad-ə-ˈkand\ *n.*

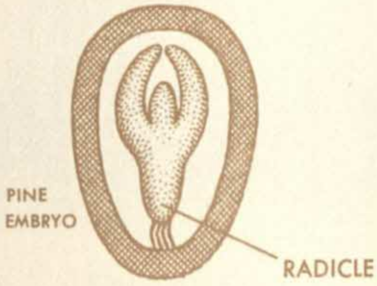
MATHEMATICS. The number or expression under a radical sign; the expression whose indicated principal root is to be found.

In the radical $\sqrt[4]{32}$, the index is 4 and the RADICAND *is 32.*

radicle \ˈrad-i-kəl\ *n.*

1. BOTANY. The early root, or root and hypocotyl, of a seed plant embryo. 2. ANATOMY. The smallest branch of a nerve or the beginning of a vein; also, a rootlike part, as the attachment of an embryo.

The edible globe of a radish forms from the hypocotyl, while the root tail forms from the RADICLE.



radioactive decay \ˌrād-ē-ōˈak-tive di-ˈkā\

CHEMISTRY and PHYSICS. The breaking up of an atomic nucleus or a free subatomic particle into two fragments as a result of forces within the nucleus or particle. One of the fragments is usually an energetic alpha or beta particle that may be accompanied by gamma radiation; see *decay series* and *half-life.*

A uranium 238 nucleus undergoes RADIOACTIVE DECAY *into a thorium nucleus and an energetic alpha particle.*

radioactive isotope \ˌrād-ē-ō-ˈak-tiv ˈī-sə-ˌtōp\

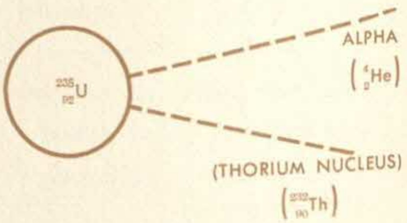
CHEMISTRY and PHYSICS. Any one of many isotopes having nuclei that break up into two fragments as a result of forces within the nucleus. Each disintegration is accompanied by the emission of alpha or beta, and usually gamma, radiation. Radioactive isotopes are also called radioisotopes; see *isotope* and *half-life.*

Every element has at least one RADIOACTIVE ISOTOPE, *although the radioactive isotope may be man-made rather than natural.*

radioactive series \ˌrād-ē-ō-ˈak-tiv ˈsi(ə)r-(ˌ)ēz\

CHEMISTRY and PHYSICS. One of four series of elements in which each succeeding member is produced by the radioactive decay of the preceding element. Each series ends in a nonradioactive element, such as lead or bismuth.

The RADIOACTIVE SERIES *that starts with uranium 238 ends with lead 206 after 13 intermediate steps.*


${}^{235}_{92}\mathrm{U}$
ALPHA
$({}^{4}_{2}\mathrm{He})$
(THORIUM NUCLEUS)
$({}^{232}_{90}\mathrm{Th})$

RADIOACTIVE DECAY

radioactivity \ˌrād-ē-ō-ˌak-ˈtiv-ət-ē\

CHEMISTRY and PHYSICS. The continuous emission of energetic subatomic particles (alpha or beta rays), or of gamma rays, from certain elements or isotopes of elements. It occurs as the result of radioactive decay; see *radioactive decay* and *curie.*

The RADIOACTIVITY *of an element is not affected by chemically combining it with a second element to form a compound.*

radio astronomy \ˈrād-ē-ˌō ə-ˈsträn-ə-mē\

ASTRONOMY. The study of celestial bodies by the radio waves they emit, as contrasted with optical astronomy in which light waves are studied. Radio astronomy may include the use of radar to study the moon, meteors and other relatively-nearby objects.

The discovery of radio stars is only one phase of RADIO ASTRONOMY *that has resulted in new evidence about the components of the universe.*

RADIO ASTRONOMY

radioautograph \ˌrād-ē-ō-ˈȯt-ə-ˌgraf\ *n.*

PHYSICS. The darkened area on a photographic film, produced by radiation from a radioactive material.

The RADIOAUTOGRAPH *is used in research to locate radioactive atoms in plants.*

radio beam \ˈrād-ē-ˌō ˈbēm\

AERONAUTICS. A navigational beam. See *beam.*

RADIO BEAM

radiocarbon dating \ˌrād-ē-ō-ˈkär-bən ˈdāt-iŋ\

CHEMISTRY and PHYSICS. Carbon dating. See *carbon dating.*

radio frequency \ˈrād-ē-ō ˈfrē-kwən-sē\

ENGINEERING and PHYSICS. Electromagnetic waves having a frequency between 10 kilocycles per second and 30 million kilocycles per second; *abbr.* RF.

A RADIO FREQUENCY *below 30 kilocycles per second is called a very low frequency (vlf).*

radioisotope \ˌrād-ē-ō-ˈī-sə-ˌtōp\ *n.*

CHEMISTRY, MEDICINE and PHYSICS. A radioactive isotope. See *radioactive isotope.*

radiology \ˌrād-ē-ˈäl-ə-jē\ *n.*

MEDICINE. The branch of medicine concerned with the use of radiant energy, especially X rays, in the diagnosis and treatment of disease or injury.

The use of the fluoroscope for examining the condition and functioning of internal organs is one technique of RADIOLOGY.

RADIOLUMINESCENCE

radioluminescence \ˌrād-ē-ō-ˌlü-mə-ˈnes-ᵊn(t)s\ *n.*

PHYSICS. The glow, or emission of light, from certain substances that have absorbed energy from a radioactive source.

The visible light of a radium-coated watch dial is an example of RADIOLUMINESCENCE.

radiometer \ˌrād-ē-ˈäm-ət-ər\ *n.*

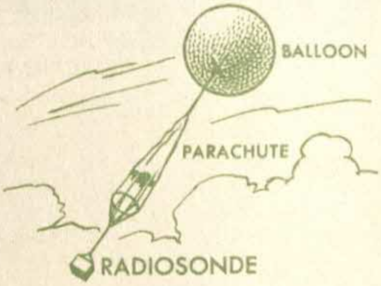
PHYSICS. A device used to detect and measure radiant heat (infrared energy). It generally has the form of a paddle wheel inside a glass bulb. The paddle wheel rotates when placed in the path of radiant energy.

The paddle wheel of a RADIOMETER *always moves so that the black surfaces of the paddles retreat from the source of energy and the silver surfaces advance.*

radiosonde \ˈrād-ē-ō-ˌsänd\ *n.*

EARTH SCIENCE. An instrument that records such conditions in the atmosphere as temperature, pressure and humidity, and that relays these readings by special radio signals.

A RADIOSONDE *is carried into the upper troposphere by a balloon and is returned to earth by a parachute when the balloon bursts.*



radio star \ˈrād-ē-ō ˈstär\

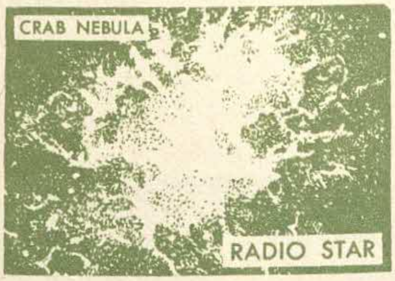
ASTRONOMY. Well-defined regions in space that are the source of radio waves that can be detected by a radio telescope; also, certain stars, such as our sun, that emit detectable radio waves.

The Crab Nebula, a strong radio source, has been classified as a RADIO STAR.

radio telescope \ˈrād-ē-ō ˈtel-ə-ˌskōp\

ASTRONOMY. An instrument that collects and focuses radio waves in much the same way that a reflecting telescope collects and focuses light waves. It consists of a radio receiver and, frequently, a movable bowl-shaped antenna.

A RADIO TELESCOPE *can be used day or night and in any kind of weather.*



radio tracking \ˈrād-ē-ō ˈtrak-iŋ\

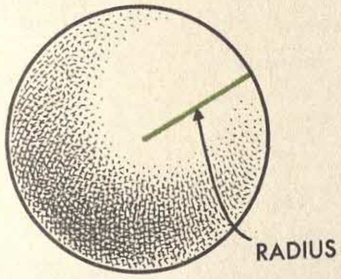
ASTRONAUTICS. The tracing of the path of satellites or other objects moving through the atmosphere or through space by radar or by radio signals emitted by the satellites or objects.

If RADIO TRACKING *is to be used to follow an object orbiting the earth, a series of radio receiving stations, located on a line encircling the earth, must be employed.*

radio wave \ˈrād-ē-ō ˈwāv\

ENGINEERING and PHYSICS. A form of radiant energy used for wireless communication. Radio waves differ from light waves by their longer wavelength and lower frequency. They are usually included within those frequencies between 10 kilocycles per second and 30 million kilocycles per second.

If an alternating electric current flows through a wire, it will give off some of its energy as a RADIO WAVE.



radium \ˈrād-ē-əm\ *n.*
CHEMISTRY. A very rare, metallic element that is highly radioactive and very active chemically. It has a half-life of about 1,622 years; see *half-life*. Symbol, Ra; atomic number, 88; atomic weight, 226.05.

A glass tube containing a small amount of RADIUM *sealed inside may explode as the result of gas pressure, since radium disintegrates into the gaseous elements radon and helium.*

radius \ˈrād-ē-əs\ *n.*
1. MATHEMATICS. A line segment extending from the center of a circle or sphere to any point on the circle or sphere. 2. ANATOMY. That one of two bones in the forearm located on the same side as the thumb.

The area of a circle equals the square of its RADIUS *times π (pi).*

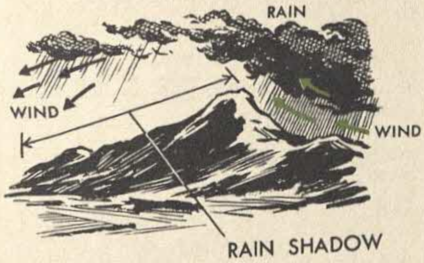
RAIN FOREST

radome \ˈrā-ˌdōm\ *n.*
ENGINEERING and PHYSICS. A dome-shaped dielectric shelter for a radar antenna.

On many commercial aircraft, the RADOME *forms the nose of the fuselage.*

rain forest \ˈrān ˈfȯr-əst\
BOTANY and EARTH SCIENCE. Any wooded area in which the annual rainfall is at least 100 inches. A rain forest is characterized by a great variety of exceedingly-tall trees and by little light penetrating to the forest floor. In tropical regions, it is called a jungle.

Nowhere on earth is there a greater variety of life forms than in a RAIN FOREST.



rain shadow \ˈrān ˈshad-(ˌ)ō\
EARTH SCIENCE. An area of land on the leeward side of a mountain or mountain range where rainfall is much less than on the windward side.

A RAIN SHADOW *may occur on any side of a mountain, depending on the prevailing winds.*

ramjet engine \ˈram-ˌjet ˈen-jən\
AERONAUTICS and ENGINEERING. A type of jet engine whose motion through the air causes air to be continuously compressed

in its open forward end. When the air is ignited with fuel in the combustion chamber, hot gases are ejected from the rear opening, providing thrust.

A RAMJET ENGINE *can be used only when it is moving fast enough to ram air into the open forward end.*

random sampling \'ran-dəm 'sam-pliŋ\

MATHEMATICS. A procedure for choosing a sample from a population so that all members of the population have an equal chance to be included in the sample.

Many polltakers use RANDOM SAMPLING *to determine public opinion on issues or to predict outcomes of political elections.*

rare-earth elements \'ra(ə)r-ˌərth 'el-ə-mənts\

CHEMISTRY. The group of chemical elements with atomic numbers 58 through 71, although the group may include the elements scandium, yttrium and lanthanum, because they have very similar chemical and physical properties; see table, page 620.

Several of the RARE-EARTH ELEMENTS *are not rare, since in the earth's crust cerium is more abundant than lead and yttrium is more abundant than tin.*

rarefaction \ˌrar-ə-'fak-shən\ *n.*

PHYSICS. A region of low air pressure, generally much lower than atmospheric pressure; in dealing with compressional waves, as of sound energy, a region of lowered pressure or density that moves at the speed of the wave, as contrasted with a compression, or a region of increased pressure or density.

As a continuous sound is transmitted through air, each RAREFACTION *is preceded and followed by compressions.*

rare gases \'ra(ə)r 'gas-əs\

CHEMISTRY. The six gaseous elements: argon, helium, krypton, neon, radon and xenon. They are frequently called inert gases because they do not tend to form compounds with other elements. All six are found in the atmosphere.

Radon, one of the RARE GASES, *is formed constantly by the radioactive decay of radium in the earth's crust.*

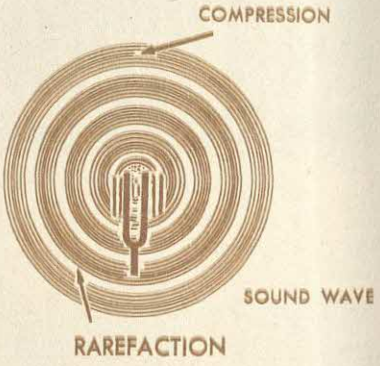
ratio \'rā-(ˌ)shō\ *n.*

MATHEMATICS. A relationship between two numbers or quantities, expressed as the quotient resulting from the division of the first number or quantity by the second number or quantity.

The RATIO *of 12 to 8 may be written as 12:8 or as 12/8 and is equal to* $\frac{3}{2}$.

Lathanum La 57	Terbium Tb 65
Cerium Ce 58	Dysprosium Dy 66
Praeseodymium Pr 59	Holmium Ho 67
Neodymium Nd 60	Erbium Er 68
Promethium Pm 61	Thulium Tm 69
Samarium Sm 62	Ytterbium Yb 70
Europium Eu 63	Lutetium Lu 71
Gadolinium Gd 64	

RARE-EARTH ELEMENTS



Element	Atomic number	Atomic weight	Electron configuration
HELIUM (He)	2	4.0026	2
NEON (Ne)	10	20.183	2-8
ARGON (Ar)	18	39.948	2-8-8
KRYPTON (Kr)	36	83.80	2-8-18-8
XENON (Xe)	54	131.30	2-8-18-18-8
RADON (Rn)	86	222.	2-8-18-32-18-8

RARE GASES

rational number \'rash-nəl 'nəm-bər\

MATHEMATICS. Any number that can be expressed as the ratio or the quotient of two integers or whole numbers.

The fraction $\frac{2}{3}$ *is a* RATIONAL NUMBER.

ray \'rā\ *n.*

1. ASTRONOMY. Any of the whitish lines that appear to radiate from lunar craters. 2. PHYSICS. A beam of radiant energy, such as light, that has a small cross section; also, a stream of particles, such as cathode rays and alpha rays. 3. BOTANY. Any of the flower stalks of a cluster of flowers called an umbel; also, any one of the marginal strap-shaped flowers of some composites, including sunflowers and asters. 4. ZOOLOGY. Any of the bony elements that support the fins of a fish; also, a radiating branch, as the arm of a starfish. 5. MATHEMATICS. A straight line that extends in one direction from a point called the origin.

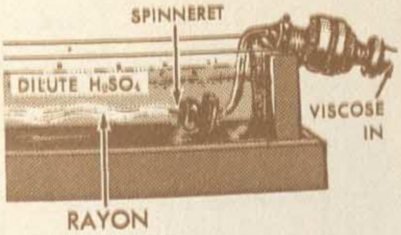
RAY

A lunar RAY *may be more than 1,000 miles long and usually originates at a large crater.*

rayon \'rā-ˌän\ *n.*

CHEMISTRY. A synthetic fiber developed as a substitute for silk and prepared by one of several processes that involve dissolving cellulose in a solvent and then precipitating it from the solution.

In making RAYON, *dissolved cellulose is forced through small holes in a plate before it is solidified as filaments in the precipitating solution.*



RAYON

reactance \rē-'ak-tən(t)s\ *n.*

PHYSICS. Opposition to the flow of alternating electric current through a circuit (as distinguished from resistance, that opposes both alternating and direct current); also, the combined effect of a coil (inductive reactance) and a capacitor (capacitive resistance) in an alternating current circuit. Reactance combines with resistance to produce the alternating current impedance of a circuit; see *impedance*.

As the frequency of alternating current increases, the inductive REACTANCE *of a circuit also increases, but its capacitive reactance decreases.*

reactants \rē-'ak-tən(t)s\ *n.*

CHEMISTRY. Substances that enter into chemical reactions, as distinguished from products that result from chemical reactions.

If baking soda is added to vinegar, the baking soda and acetic acid in the vinegar are REACTANTS *that produce fizzing.*

reaction \rē-'ak-shən\ *n.*

1. CHEMISTRY. A process by which at least one new substance is produced from a beginning substance or substances (reactants), with accompanying release or using up of energy; a chemical change. 2. PHYSICS. The equal and opposite force that accompanies the application of any initial force, as the recoil of a gun; also, any change in the nucleus of an atom. 3. MEDICINE and PHYSIOLOGY. The response of a living thing to a stimulus; also, a variety of chemical changes that may occur in the body as a result of the presence of some substance, such as pollen, not normally present.

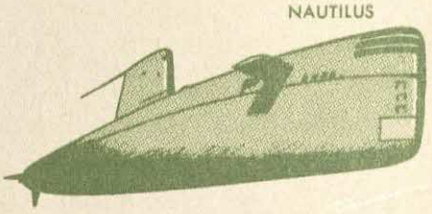
The burning of paper is a chemical REACTION *in which paper and oxygen are the reactants and carbon dioxide, water vapor and ash are the products.*

REACTION

reaction time \rē-'ak-shən 'tīm\

PHYSIOLOGY. The period of time between the application of a stimulus and the start of the response to the stimulus; also called latent period.

The REACTION TIME *required to apply automobile brakes varies with different people.*



REACTOR

reactor \rē-'ak-tər\ *n.*

1. ENGINEERING. A device in which fissionable materials undergo a chain reaction at a controlled rate; see *nuclear reactor.* 2. CHEMISTRY. A piece of equipment in which a chemical reaction is carried out. 3. MEDICINE. An animal that reacts positively to a test, as a test for tuberculosis.

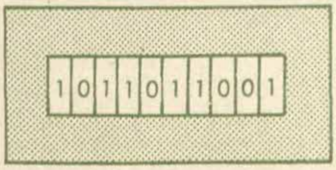
The first ship to be powered by a nuclear REACTOR *was the submarine "Nautilus."*

readout \'rē-ˌdau̇t\ *n.*

ENGINEERING and MATHEMATICS. A readable form of data produced by a computer.

Many analog computers produce READOUT *in the form of printed numbers.*

This 10-digit binary word signals Surveyor moon probe to move its antenna



READOUT

ON DEVICE TRANSMITTING COMMANDS TO SPACECRAFT

reagent \rē-'ā-jənt\ *n.*

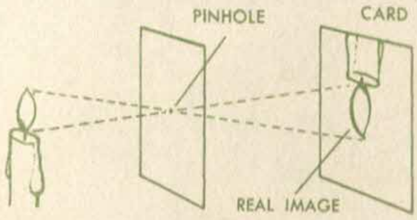
CHEMISTRY. Any chemical used to react with another substance. A reagent is frequently in the form of a solution; also, a substance of high purity used in chemical analysis.

Hydrogen sulfide is often used as a REAGENT *to detect such metallic ions as lead and cadmium in an unknown solution.*

real image \'rē(-ə)l 'im-ij\

PHYSICS. An inverted image that may be either larger or smaller than the object. It is formed by rays of light that have passed through a lens or that have been reflected from a curved mirror and can be focused on a sheet of paper or a screen; see *image* and *virtual image.*

A REAL IMAGE *can be formed by allowing light from an object to pass through a pinhole and fall on a screen.*



REAL IMAGE

real number \'rē(-ə)l 'nəm-bər\

MATHEMATICS. Any rational number, such as a whole number or fraction, or any irrational number, such as a nonrepeating decimal.

The irrational number $\sqrt{3}$ *is a* REAL NUMBER.

recapitulation theory \ˌrē-kə-ˌpich-ə-'lā-shən 'thē-ə-rē\

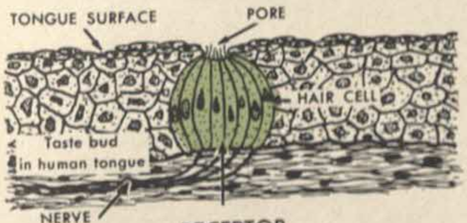
BIOLOGY. A theory maintaining that the phases in the embryonic development of an animal are identical with (or recapitulate) the evolutionary changes of its adult ancestors; also called the biogenetic law; see *ontogeny* and *phylogeny.*

Biologists now regard a living organism as one individual from zygote to death, not as a series of ancestors as originally stated in the RECAPITULATION THEORY.

receptor \ri-'sep-tər\ *n.*

PHYSIOLOGY. A sense organ, such as an eye; also, a sensory neuron or a single cell that functions as a receiver of a specific stimulus.

Each RECEPTOR *is usually adapted to receiving only one kind of stimulus, such as light, odor, pressure or heat.*



RECEPTOR

recessive trait \ri-'ses-iv 'trāt\

BIOLOGY. A hereditary characteristic that does not appear when combined with a dominant characteristic.

Albinism, or lack of pigmentation, is a common RECESSIVE TRAIT *that is frequently masked by the dominant trait of pigmentation.*

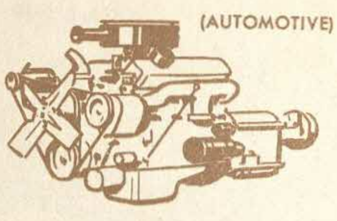
reciprocal \ri-'sip-rə-kəl\

MATHEMATICS (N.). The result obtained when 1 is divided by a given number or quantity; also, the multiplicative inverse of an expression. (Adj.). Related or corresponding, but inverse in nature.

The RECIPROCAL *of 9 is 1/9.*

reciprocating engine \ri-'sip-rə-ˌkāt-iŋ 'en-jən\
ENGINEERING. An engine in which the reciprocating, or back-and-forth, motion of a piston is changed to rotary motion by a crankshaft and connecting rod.

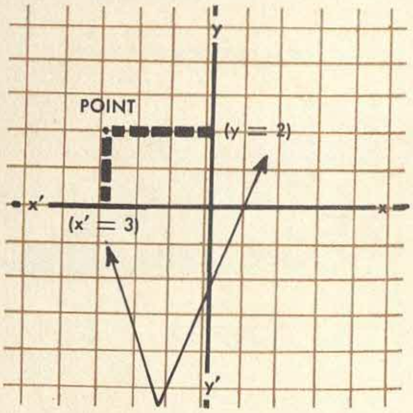
The RECIPROCATING ENGINE *of the automobile may eventually be replaced by a turbine.*



RECIPROCATING ENGINE

rectangular coordinates \rek-'taŋ-gyə-lər kō-'ȯrd-nəts\
MATHEMATICS. The two numbers, x and y, of an ordered pair (x,y) by which a point may be located in a plane divided into four quadrants by a horizontal line, or x-axis, and by a vertical line, or y-axis. The first number of the pair indicates the number of units the point is above or below the x-axis, while the second number of the pair indicates the number of units the point is to the right or left of the y-axis; also called Cartesian coordinates.

The RECTANGULAR COORDINATES *of a point are often referred to individually as the abscissa and the ordinate.*



RECTANGULAR
COORDINATES

rectification \ˌrek-tə-fə-'kā-shən\ *n.*
1. MATHEMATICS. The process of determining a straight line segment whose length is equal to that of a given curve or arc of that curve. 2. PHYSICS. A process that changes alternating electrical current into direct current; see *rectifier.*

The RECTIFICATION *of many curves may be achieved by integral calculus.*

rectifier \'rek-tə-ˌfī(-ə)r\ *n.*
PHYSICS. A device that permits electric current to flow through it in only one direction. It is used to change alternating current into direct current and may be an electron tube, a solid state device, such as a germanium crystal, or an electrolytic cell in which two electrodes are immersed in a conducting solution.

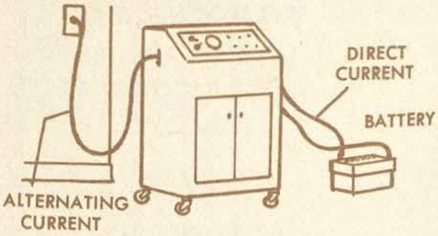
A RECTIFIER *is necessary to charge a storage battery by means of alternating electric current.*

rectilinear \ˌrek-tə-'lin-ē-ər\ *adj.*
MATHEMATICS. Referring to straight lines; composed of, formed by or bounded by straight lines.

An object moving along a straight line has RECTILINEAR *motion.*

rectilinear motion \ˌrek-tə-'lin-ē-ər 'mō-shən\
PHYSICS. Movement in a straight line.

What appears to be RECTILINEAR MOTION *may actually be curved motion, as a car moving across the curved surface of the earth.*



RECTIFIER (BATTERY CHARGER)

rectum \\'rek-təm\\ *n.*

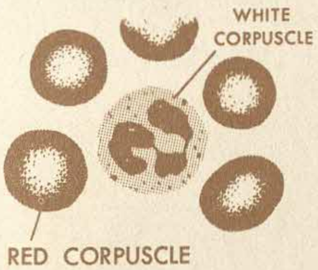
ANATOMY and ZOOLOGY. The terminal part of the alimentary canal, or food tube, in an animal.

A grasshopper's intestine expands at its lower end to form the RECTUM.

recycling \\ˌrē-'sī-kliŋ\\ *n.*

ASTRONAUTICS. The recovery, purification and reuse of waste materials, particularly of air, food and water, in a space vehicle.

One plan for RECYCLING *involves the use of green plants to convert carbon dioxide and solid wastes to oxygen and edible carbohydrates.*



red corpuscle \\'red 'kȯr-ˌpəs-əl\\

ANATOMY and ZOOLOGY. A disk-shaped, double-concave blood cell in vertebrate animals. It contains hemoglobin that carries oxygen from the lungs or gills to the tissues; also called an erythrocyte.

A single RED CORPUSCLE *appears yellowish through a microscope.*

redox \\'rē-ˌdäks\\ *adj.*

CHEMISTRY. Referring to a type of chemical reaction in which one atom or group of atoms gains electrons (is reduced) and another atom or group of atoms loses electrons (is oxidized); also called oxidation-reduction.

Hydrogen is widely used as a reducing agent in industrial REDOX *reactions.*

RUSTING OF IRON

$4Fe + 3O_2 + 6H_2O \rightarrow 2(Fe_2O_3 \cdot 3H_2O)$

TARNISH OF SILVER

$2Ag + S \rightarrow Ag_2S$

REDOX
(REACTIONS)

red shift \\'red 'shift\\

ASTRONOMY. A shift in the lines of certain stellar and galactic spectra toward the red end because those stars or galaxies are moving away from the earth; see *Doppler shift* and *Doppler effect.*

The RED SHIFT *in the spectra of distant galaxies supports the theory of the expanding universe.*

reducing agent \\ri-'d(y)üs-iŋ 'ā-jənt\\

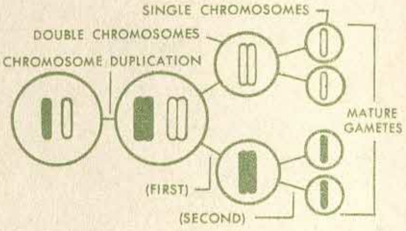
CHEMISTRY. An atom, group of atoms or an ion that provides, or loses, electrons that are gained by another atom, group of atoms or an ion in a chemical reaction; also, a substance that will remove oxygen from a compound.

Hydrogen is a REDUCING AGENT *that reacts with hot iron oxide to produce water vapor and metallic iron.*

reduction \ri-'dək-shən\ *n.*

1. MATHEMATICS. The process of changing an expression or quantity to a different, usually simpler, form by such procedures as making substitutions, collecting terms and simplifying fractions; in geometry, proceeding from one proposition to another that seems simpler, proving the latter and then reversing the order; also, a decrease. 2. CHEMISTRY. A chemical reaction by which an atom or ion gains one or more electrons; also, the removal of oxygen from a compound, usually forming a free metallic element as one product. 3. ENGINEERING. The removal of nonmetallic impurities from ores to produce metals.

The REDUCTION *of a fraction to a decimal is achieved by dividing the numerator by the denominator.*



REDUCTION DIVISION

reduction division \ri-'dək-shən də-'vizh-ən\

BIOLOGY. The first of the two cell divisions in meiosis, in which the chromosome number in a cell is reduced by half. Reduction division occurs only in gametogenesis, or meiosis, and leads to the haploid number of chromosomes in the reproductive cells of plants and animals; see *meiosis, haploid number* and *diploid number.*

A sperm and an egg formed by REDUCTION DIVISION *unite in fertilization to restore a full number of chromosomes in the fertilized cell.*

REEF (ATOLL)

reef \'rēf\ *n.*

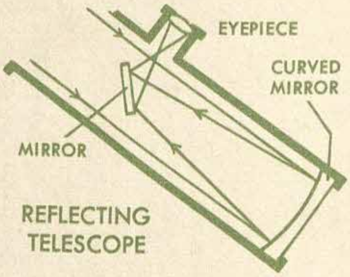
EARTH SCIENCE. A ridge or range of rock or coral at or near the surface of the sea. It is generally submerged but may be visible at low tide.

An atoll, found typically in the South Pacific Ocean, is a coral REEF *or ring-shaped chain of coral reefs.*

reflecting telescope \ri-'flekt-iŋ 'tel-ə-ˌskōp\

ASTRONOMY. A type of telescope that uses a spherical or parabolic mirror rather than an objective lens for focusing light rays; also called a reflector; see *refracting telescope* and *Cassegrainian reflector.*

The largest REFLECTING TELESCOPE *in the world, at the Palomar Observatory in California, has a mirror 200 inches in diameter.*



REFLECTING TELESCOPE

reflection \ri-'flek-shən\ *n.*

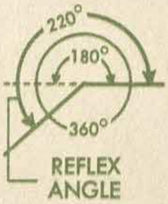
PHYSICS. The change of direction of a moving object, a stream of particles or energy when it strikes a substance that it does not enter; also, the image formed by a mirror.

A flat stone, when thrown properly onto a still lake or pond, will undergo repeated REFLECTION.

reflex \'rē-ˌfleks\

1. PHYSIOLOGY (N.). A simple involuntary response to a stimulus, often involving only three neurons: a receptor, an adjustor and an effector. 2. PHYSICS (Adj.). Referring to an electronic amplifier that can amplify simultaneously two signals of widely-different frequencies.

The blinking of an eye that is threatened by a moving object is a simple REFLEX.



reflex angle \'rē-ˌfleks 'aŋ-gəl\

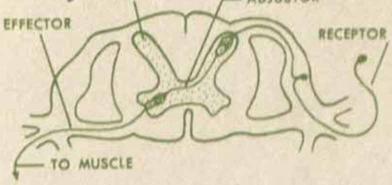
MATHEMATICS. An angle that is greater than 180 degrees and less than 360 degrees.

An angle that measures 220 degrees is a REFLEX ANGLE.

reflex arc \'re-ˌfleks 'ärk\

PHYSIOLOGY. The nerve path or route traveled by an impulse in an involuntary response to a stimulus. The path often consists simply of a receptor, an adjustor and an effector neuron. The receptor nerve receives the stimulus and transfers the impulse to an adjustor neuron in the central nervous system. The adjustor neuron transmits the impulse to an appropriate motor or effector neuron, which, in turn, stimulates a muscle or a gland to act.

Many of the adjustor neurons that function in a REFLEX ARC *are located within the spinal cord.*



REFLEX ARC

reflex camera \'rē-ˌfleks 'kam-(ə-)rə\

ENGINEERING. A camera that uses a mirror or prism system in the viewfinder. It usually forms an image the same size as the image that is formed on the photographic film.

A single lens REFLEX CAMERA *forms both the viewfinder image and the image on the film with light coming through the lens, while a double lens reflex camera has a separate viewfinding lens that is usually a duplicate of the main lens.*

reflux \'rē-ˌfləks\ *v.*

CHEMISTRY. To heat a solution and condense the vapors formed into a liquid that flows back into the solution to be heated again.

In carrying out slow reactions of highly-volatile compounds, it is often desirable to REFLUX *the reactants so the volatile materials will not be lost.*